

# Integrierte Benutzerunterstützung für die Visualisierung in Geo-Informationssystemen

Vom Fachbereich Informatik  
der Technischen Universität Darmstadt  
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades eines  
Doktor-Ingenieurs (Dr.-Ing.)  
von

Dipl.-Inform. Volker Jung  
aus Alzey

Referenten der Arbeit:	Prof. Dr. J. L. Encarnação Prof. Dr. H. Schumann
------------------------	---

Tag der Einreichung:	17. Juli 1998
Tag der mündlichen Prüfung:	28. August 1998

D 17

Darmstädter Dissertation 1998



## Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Fachgebiet Graphisch-Interaktive Systeme im Fachbereich Informatik der Technischen Universität Darmstadt und am Fraunhofer-Institut für Graphische Datenverarbeitung in Darmstadt.

Mein besonderer Dank gilt Herrn Prof. Dr. J.-L. Encarnação, der mir die Möglichkeit gab, an seinen Instituten zu arbeiten und der durch vielfältige Unterstützung meiner Forschungstätigkeit zum Gelingen dieser Arbeit beitrug.

Frau Prof. Dr. H. Schumann danke ich herzlich für das frühzeitige Interesse an dieser Arbeit, für die Übernahme des Korreferats und für die schnelle, gründliche und wissenschaftlich kritische Durchsicht des Manuskripts.

Von den vielen Fachkollegen, mit denen ich meine Arbeit diskutiert und die mir wertvolle Anregungen gegeben haben, möchte ich vier besonders erwähnen. Herrn Dipl.-Inform. Wolfgang Müller danke ich für frühe, einflußreiche Diskussionen, wichtige Literaturhinweise und stetige Ermunterung, Herrn Prof. Ernesto Arias, Ph.D., für seinen Enthusiasmus und seinen kritischen und menschlichen Blick, Frau Dr. Natalia Andrienko und Herrn Dr. Gennady Andrienko für den interessanten und fruchtbaren wissenschaftlichen Austausch.

Mein Dank gilt allen Mitarbeiterinnen und Mitarbeitern der Abteilung Graphische Informationssysteme am Fraunhofer-IGD für die freundliche und kreative Arbeitsatmosphäre und die kollegiale Hilfe bei den kleinen und großen Problemen der täglichen Arbeit. Besonders danke ich meiner langjährigen Abteilungsleiterin, Frau Dr. Christine Giger, die meine wissenschaftliche Arbeit vom ersten Tag an gefördert hat, mir den nötigen Freiraum verschafft hat und zu jeder Zeit eine wertvolle Ratgeberin war.

Ich möchte mich auch bei allen Studentinnen und Studenten bedanken, die im Rahmen von Studien- und Diplomarbeiten unter meiner Betreuung ebenfalls zum Gelingen dieser Arbeit beigetragen haben. Dies gilt insbesondere für Stefan Göbel, Michael Harkort, Markus Kühnel, Ana Mafalda Martins, Friedhelm Schanz und Jörg Schmidt.

Meiner Frau, Dipl.-Math. Christine Schanz, danke ich ganz besonders für ihr Verständnis und ihre Unterstützung in allen Phasen dieser Arbeit.

Volker Jung

Darmstadt, im September 1998



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
1.1	Geo-Informationssysteme	1
1.2	Visualisierung in Geo-Informationssystemen	2
1.3	Problemstellung	3
1.4	Ziel und Aufbau der Arbeit	5
<b>2</b>	<b>Die Problematik: Effektivität und Nutzbarkeit der GIS-Visualisierung</b>	<b>7</b>
2.1	Geo-Informationssysteme: Stand der Technik	7
2.1.1	Aufbau und Operationen eines GIS	8
2.1.2	GIS-Daten und -Datenmodelle	10
2.1.3	Wichtige GIS-Anwendungsgebiete	12
2.2	Computervisualisierung: eine neue wissenschaftliche Fachrichtung	15
2.2.1	Formen und Anwendungsgebiete der Computervisualisierung	16
2.2.2	Wissensquellen der Computervisualisierung	18
2.3	Der Begriff „GIS-Visualisierung“	18
2.4	Techniken der GIS-Visualisierung	19
2.4.1	Thematisch-kartographische Grundbegriffe	20
2.4.2	Choroplethen	21
2.4.3	Kreissignaturen	22
2.4.4	Weitere Formen thematischer Karten	23
2.4.5	Darstellungen von Multiparameterdaten	26
2.4.6	Diagramme und Diagrammkarten	27
2.4.7	Dreidimensionale Darstellungen	28
2.4.8	Visualisierung von Vektorfeldern	29
2.4.9	Animation	31
2.5	Effektive GIS-Visualisierung	32
2.5.1	Ausdrucksfähigkeit und Effektivität von Visualisierungen	32
2.5.2	Elemente der Graphischen Semiotik	34
2.5.3	Der Entwurf thematischer Karten	36
2.5.4	Methodik der praktischen Computervisualisierung	37
2.6	Das Problemfeld GIS-Visualisierungsentwurf	38
2.6.1	Wissenslücken bei der GIS-Visualisierung	38
2.6.2	Häufige Fehler	39
2.6.3	Lösungsansätze	41
2.7	Der Lösungsweg: eine Umgebung zur Benutzerunterstützung	42

<b>3</b>	<b>Automatisierter Visualisierungsentwurf: Stand der Technik</b>	<b>45</b>
3.1	Charakteristika automatisierter Präsentationssysteme	45
3.2	Automatisierter Visualisierungsentwurf: Zwei Fallstudien	47
3.2.1	APT: Automatisierte Präsentation relationaler Informationen	47
3.2.2	VISTA: Wissensbasiertes System für den Visualisierungsentwurf	51
3.3	Charakterisierungsschemas für den Visualisierungsentwurf	53
3.3.1	Datencharakteristika	54
3.3.2	Visualisierungszielsetzung	55
3.3.3	Benutzermodell	58
3.4	Automatisierungsansätze in der computergestützten Kartographie	59
3.4.1	Der kartographische Entwurfsprozeß	60
3.4.2	Automatisierte Kartenbeschriftung	62
3.4.3	Architekturen kartographischer Expertensysteme	63
3.4.4	Wissensbasierte Symbolisierung: zwei Prototypsysteme	65
3.5	Automatisierter GIS-Visualisierungsentwurf: Offene Probleme	67
<b>4</b>	<b>Charakterisierung von Sachdaten und GIS-Visualisierungskontext</b>	<b>69</b>
4.1	Charakteristika von GIS-Sachdaten	69
4.1.1	Geographisches Bezugssystem	70
4.1.2	Relationen und Variablen	71
4.1.3	Geographische Variablen	73
4.1.4	Attributvariablen	74
4.1.5	Datenqualität	78
4.1.6	Ein formales Charakterisierungsschema	81
4.1.7	Quellen für die Datencharakterisierung	86
4.2	Visualisierungszielsetzung	87
4.2.1	Funktion der Visualisierung	88
4.2.2	Aufgabe der Visualisierung	89
4.2.3	Prioritäten	91
4.2.4	Ein formales Charakterisierungsschema	91
4.3	Visualisierungskontext	92
4.3.1	Ressourcenmodell	92
4.3.2	Benutzermodell	94
4.3.3	Ein formales Kontextmodell	96
4.3.4	Quellen des Kontextmodells	98
4.4	Zusammenfassung und Diskussion	99
<b>5</b>	<b>Eine Methodik für den automatisierten GIS-Visualisierungsentwurf</b>	<b>101</b>
5.1	Repräsentation graphischen Wissens	102
5.1.1	Klassen graphischer Regeln	102
5.1.2	Ausdrucksfähigkeit	103
5.1.3	Effektivität	103
5.1.4	Ausgestaltung und Parameterwahl	105
5.2	Eine Strategie für den GIS-Visualisierungsentwurf	106

5.3	Die Entwurfsmethodik im Überblick .....	107
5.4	Primitiver Visualisierungsentwurf .....	109
5.4.1	Raumbezogene Datensätze .....	109
5.4.2	Dekomposition von Relationen .....	110
5.4.3	Matching und Primitiver Entwurf .....	112
5.4.4	Regeln zur Ausdrucksfähigkeit .....	113
5.5	Komposition von Visualisierungen .....	115
5.5.1	Überlagerung .....	116
5.5.2	Sequenz .....	117
5.5.3	Iteration .....	118
5.5.4	Animation .....	119
5.5.5	Sukzessive Komposition .....	120
5.6	Parameteroptimierung .....	121
5.6.1	Parametertypen .....	121
5.6.2	Parameterregeln .....	123
5.6.3	Verwaltung visueller Ressourcen .....	124
5.6.4	Datenklassifikation .....	126
5.7	Effektivitätsabschätzung .....	126
5.7.1	Fuzzy-Mengen .....	127
5.7.2	Wie läßt sich Effektivität repräsentieren? .....	130
5.7.3	Syntax und Fuzzy-Repräsentation der Effektivität .....	131
5.7.4	Rechnen mit Effektivitäten .....	132
5.7.5	Wie läßt sich Effektivität abschätzen? .....	134
5.7.6	Primitive Effektivitätsabschätzung .....	135
5.7.7	Komplexe Effektivitätsabschätzung .....	140
5.7.8	Defuzzifizierung .....	142
5.8	Erweiterungen der Methodik .....	145
5.8.1	Kompakte Visualisierung mehrerer Relationen .....	145
5.8.2	Einschränkung des Suchbaums .....	146
5.9	Zusammenfassung und Diskussion .....	147
<b>6</b>	<b>Empirische Bestimmung der Effektivität von GIS-Visualisierungsformen</b>	<b>151</b>
6.1	Konzeption der Effektivitätstests .....	152
6.1.1	Hypothesen .....	152
6.1.2	Testdesign .....	152
6.2	Testaufbau und Umsetzung .....	154
6.2.1	Testaufbau zu Werte Ablesen .....	156
6.2.2	Testaufbau zu Werte Finden .....	159
6.2.3	Testaufbau zu Werte Vergleichen .....	160
6.2.4	Testaufbau zu Trend Erkennen .....	161

6.3	Statistische Auswertung .....	162
6.3.1	Werte Ablesen .....	163
6.3.2	Werte Finden .....	163
6.3.3	Werte Vergleichen .....	164
6.3.4	Trend Erkennen .....	165
6.4	Interpretation .....	165
6.4.1	Werte Ablesen .....	165
6.4.2	Werte Finden .....	166
6.4.3	Werte Vergleichen .....	167
6.4.4	Trend Erkennen .....	167
6.4.5	Die zentralen Hypothesen .....	167
6.4.6	Diskussion .....	168
<b>7</b>	<b>VIZARD: GIS-Visualisierungsumgebung mit Benutzerunterstützung</b>	<b>171</b>
7.1	VIZARD .....	171
7.2	Regelbasis .....	173
7.2.1	Ausdrucksfähigkeit .....	173
7.2.2	Primitive Effektivität .....	174
7.2.3	Komplexe Effektivität .....	176
7.3	Benutzerschnittstelle und Beispielanwendungen .....	177
7.3.1	Spezifikation von Datensatz und Kontext .....	177
7.3.2	Entwurf und Präsentation von Visualisierungen .....	180
7.3.3	Erklärung von Visualisierungen .....	185
7.4	Zusammenfassung und Bewertung .....	187
<b>8</b>	<b>Zusammenfassung und Ausblick</b>	<b>189</b>
8.1	Zusammenfassung der Ergebnisse .....	189
8.2	Diskussion .....	192
8.3	Ausblick .....	194
8.3.1	Weiterführende Forschungsarbeiten .....	194
8.3.2	CommonGIS .....	195
	<b>Glossar</b>	<b>197</b>
	<b>Literaturverzeichnis</b>	<b>201</b>



# Kapitel 1

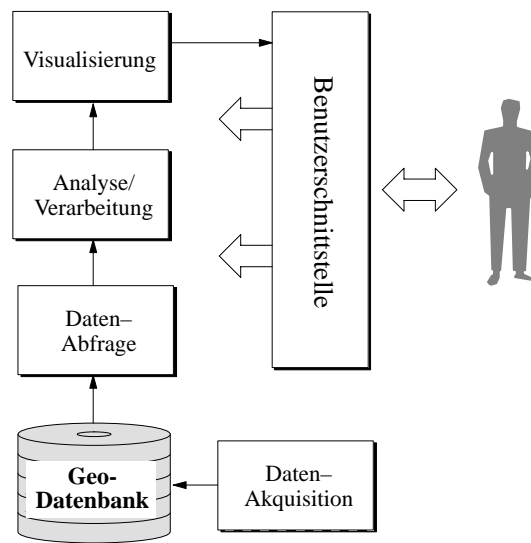
## Einführung

### 1.1 Geo-Informationssysteme

Geo-Informationssysteme (GIS) sind ein verbreitetes Werkzeug für die Verwaltung und die Analyse raumbezogener Information. Der GIS-Boom, in dem wir uns gerade befinden, wurde durch eine neue Qualität der Anwendungen ausgelöst: Das Potential von GIS wurde erkannt und folgerichtig gehen die GIS-Anwendungsbereiche mittlerweile erheblich über die Erfassung und Verwaltung raumbezogener Daten hinaus. GIS werden heute verstärkt als Werkzeuge für die raumbezogene Planung, Entscheidungsunterstützung und Analyse räumlicher Prozesse eingesetzt.

GIS zählen zu den komplexesten Informationssystemen, die auf dem Markt angeboten werden. Sie verfügen häufig über eine sehr umfassende und manchmal schwer überschaubare Funktionalität, so hat beispielsweise ein weit verbreitetes, kommerzielles GIS ca. 1800 verschiedene Funktionen [Fran93]. Die Module der GIS-Software bauen, wie in allen Informationssystemen, sehr stark auf einer Datenbankschnittstelle auf. Abb. 1-1 zeigt den modularen Aufbau eines typischen Geo-Informationssystems. Im Zentrum steht eine Geodatenbank, die raumbezogene Daten effektiv speichern kann. Auch die Datenzugriffsmethoden sind für Geodaten optimiert und lassen Abfragekriterien wie räumliche Lage und räumliche Nachbarschaft zu. Eine spezielle Datenakquisitionskomponente ermöglicht die Eingabe und Aktualisierung des Geodatenbestandes. Datensätze können im GIS mit besonderen Analyse- und Verarbeitungsfunktionen ausgewertet und kombiniert werden. Die Visualisierungs- und Präsentationskomponente stellt die Geodatenätze oder die Analyseergebnisse schließlich in graphischer Form dar.

Nach A. U. Frank sind es vor allem drei gesellschaftliche Faktoren, die die Entwicklung der GIS vorantreiben und den Bedarf nach GIS noch erhöhen werden [Fran93]: (1) das steigende Umweltbewußtsein, (2) der wachsende Informationsbedarf der Öffentlichkeit und (3) die Notwendigkeit der Kostensenkung in der öffentlichen Verwaltung. Durch die globalen Datennetze und das zunehmende Vordringen der Informations- und Kommunikationstechnik in den privaten Bereich, werden die Anwendungsbereiche von GIS aller Wahrscheinlichkeit nach noch zunehmen. Jean-Claude Muller gab 1993 die Schätzung ab, im Jahr 2000 würden mindestens eine Million Anwender regelmäßig GIS benutzen



**Abb. 1-1:** Architektur eines Geo-Informationssystems

[Mull93], bei anhaltendem Ausbau der Informationsinfrastruktur rechnet er mit einer sogar wesentlich höheren Zahl.

## 1.2 Visualisierung in Geo-Informationssystemen

Die Visualisierung nimmt im Bereich der Geo-Informationssysteme eine Schlüsselrolle ein. Man versteht unter Visualisierung ganz allgemein das Umwandeln symbolischer oder numerischer Information in visuelle Information wie Bilder, Graphiken und Karten. Visualisierung ermöglicht GIS-Anwendern das Erfassen komplexer, räumlicher Zusammenhänge und ist deswegen eine wichtige Voraussetzung für die korrekte Analyse einer Planungssituation und für das Treffen fundierter räumlicher Entscheidungen, wie z.B. bei der Standortwahl. Auch in der Präsentation und Kommunikation solcher räumlicher Entscheidungen und ihrer Hintergründe stellt die Visualisierung ein wichtiges Hilfsmittel dar.

Mit der Einführung der computergestützten Kartographie und der GIS ab dem Ende der siebziger Jahre fand auch eine Erweiterung der kartographischen Techniken statt. Viele der Darstellungsmethoden, die bisher als aufwendig galten und deswegen wenig Verwendung fanden, waren jetzt mit Computerunterstützung leicht realisierbar. Ab dem Ende der Achtziger Jahre fanden dann dreidimensionale Darstellungen und die Techniken der technisch-wissenschaftlichen Visualisierung (Visualization in Scientific Computing, ViSC) in Geo-Informationssystemen Einzug. Gründe dafür waren unter anderem die enormen Datenmengen, die durch moderne Fernerkundungsmethoden und Satellitenbeobachtung gewonnen werden konnten. Neuartige Anwendungen, insbesondere im Umweltbereich und der Simulation, arbeiteten mit Vektorfeldern und 3D-Daten und erforderten deswegen spezielle Visualisierungsformen.

Ende der neunziger erweitern zwei neue Einflüsse die Techniken und Anwendungsgebiete der GIS-Visualisierung. Einerseits bringt die Entwicklung der Virtuellen Realität neue Dimensionen von Interaktivität in alle graphischen Anwendungen, also auch in die Computervisualisierung und in die GIS. Andererseits befinden wir uns mit der Entwicklung der globalen Informationsinfrastruktur in einer zweiten Computerrevolution: sehr viele Menschen werden Zugang zu raumbezogenen Informationen erlangen und diese Möglichkeit auch nutzen wollen. Hier werden thematische Karten weiterhin eine große Rolle als anschauliche und vertraute Darstellungsform behalten. Sie müssen aber auf das neue Medium angepaßt werden, denn Bildschirmdarstellungen haben in ihrer Dynamik andere Charakteristika als beispielsweise Druckerzeugnisse.

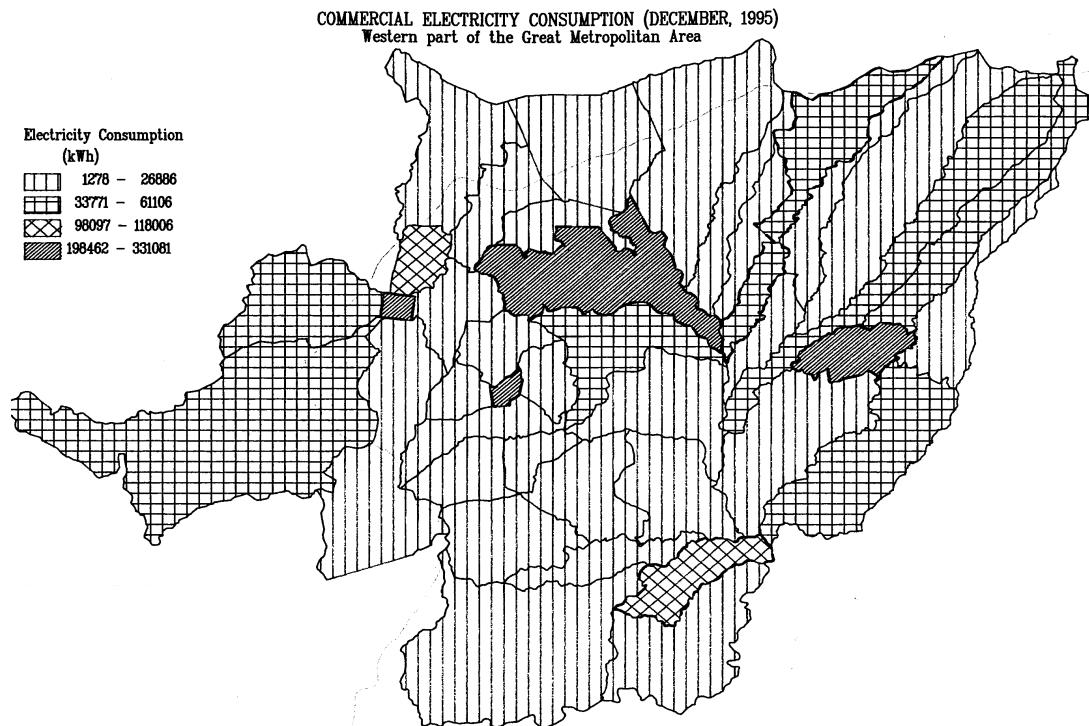
### 1.3 Problemstellung

Insgesamt läßt sich also sagen, daß die Computerkartographie immer mehr von einer umfassenderen GIS-Visualisierung abgelöst wird. Heutige GIS stellen eine große Zahl von Visualisierungstechniken zur Verfügung. Neben den Methoden der computergestützten thematischen Kartographie werden teilweise schon aufwendige, dreidimensionale Visualisierungen unterstützt.

Für die Erstellung aussagekräftiger und effektiver GIS-Visualisierungen ist jedoch die Funktionalität des Visualisierungssystems nicht alleine ausschlaggebend: Visualisieren bedeutet immer auch Entwerfen, ist folglich ein kreativer und wissensintensiver Prozeß. Auf dem Gebiet der GIS-Visualisierung sind Kenntnisse des graphischen Entwurfs, der Kartographie und der Computervisualisierung erforderlich. Modernen GIS-Visualisierungssystemen fehlt jedoch dieses Wissen, das doch eigentlich zu den Visualisierungstechniken gehört und für deren sinnvollen Einsatz notwendig ist. Die GIS-Visualisierungskomponente wurde deshalb sehr treffend als funktionsreich und wissensarm charakterisiert [YaST93].

In der Anfangszeit von GIS hatte noch die Mehrzahl der GIS-Anwender eine fundierte kartographische Ausbildung. Heute, mit der Verbreitung von GIS in den verschiedensten Anwendungsbereichen, stammen viele der GIS-Anwender aus Fachgebieten, wo die Kartographie nicht Bestandteil der Ausbildung ist. Sie haben sich kartographische Kenntnisse durch Beobachtung und Experimente angeeignet und kennen häufig fundamentale Gesetze der graphischen Kommunikation nicht. Trotzdem ermöglichen die GIS es ihnen, eindrucksvolle Visualisierungen und Karten zu erstellen. Diese Visualisierungen sehen zwar sehr professionell aus, erfüllen aber häufig ihren Zweck, die effektive Kommunikation raumbezogener Information, nur unzureichend. Dies kann sogar soweit gehen, daß die ursprüngliche Information in der Darstellung vollkommen verfälscht wiedergegeben wird.

Ein Beispiel für eine fehlerhafte kartographische Darstellung ist in Abb. 1-2 zu sehen. In dieser von einer Stadtplanungsagentur erstellten Karte soll der Stromverbrauch verschiedener Bezirke einer Region gegenübergestellt werden. Die gewählte Darstellungsart, eine sogenannte Choroplethenkarte, schraffiert jeden Bezirk in der Karte entsprechend dem Gesamtverbrauch während eines Monats. Sie ist jedoch in zweifacher Hinsicht ungeeig-



**Abb. 1-2:** Fehlerhafte kartographische Darstellung

net, diesen Datensatz darzustellen. Zum einen entsteht durch die flächenhafte Schraffur der Eindruck, es seien nicht absolute Werte (Gesamtverbrauch) sondern relative Werte (Verbrauch pro Flächeneinheit) dargestellt. Ein falscher Eindruck entsteht besonders dann, wenn die Größe der Bezirke sehr unterschiedlich ist, was in diesem Beispiel auch der Fall ist. Bevor diese absoluten Verbrauchszahlen mit Choroplethen dargestellt werden dürfen, müssen sie erst durch den Flächeninhalt des jeweiligen Bezirks dividiert und dadurch normiert werden.

Ein zweites Problem dieser Karte ist die Repräsentation numerischer Werte durch verschiedene Arten von Schraffuren. Die Zuordnung von Schraffurart zu Wertebereich ist zwar eindeutig, aber nicht offensichtlich. Will der Betrachter z.B. zwei Bezirke vergleichen, muß in der Regel in der Legende nachgesehen werden, welche Schraffurart dem höheren Wertebereich entspricht.

Wie kann die Lücke graphischen Wissens in GIS-Anwendungen geschlossen und fehlerhafte GIS-Visualisierung vermieden werden? Es gibt zwei verschiedene Lösungsansätze für dieses Problem. Zum einen kann man versuchen, die Wissenslücke auf der Anwenderseite zu schließen und möglichst viele potentielle Benutzer von GIS in graphischem Entwurf und in der graphischen Kommunikation von Information ausbilden. Die Alternative dazu ist, die Wissenslücke auf der Systemseite zu schließen und die GIS-Visualisierungskomponente mit graphischem Wissen ausstatten. Beide Lösungsansätze bringen jedoch Schwierigkeiten mit sich. Eine vollständige Automatisierung des GIS-Visualisierungsentwurfs scheint wegen des Umfangs und der Komplexität des erforderlichen Wissens nicht

realistisch. Eine Ausbildung aller potentiellen Anwender von GIS-Visualisierung kann ebenfalls nicht garantiert werden. Zur Lösung dieser Problematik bietet sich jedoch eine Synthese beider Ansätze an: Graphisches Wissen sollte einerseits formalisiert werden, um den GIS-Benutzer beim Entwurf von Visualisierungen direkt zu unterstützen und den Entwurf teilweise oder vollständig zu automatisieren. Andererseits sollte das graphische Wissen aber auch in einer den Anwendern zugänglichen Form abgelegt sein, um ihnen bei Bedarf die Möglichkeit einer gezielten Weiterbildung auf dem Gebiet des Visualisierungsentwurfs zu geben.

## 1.4 Ziel und Aufbau der Arbeit

Zusammenfassend liegt das Ziel dieser Forschungsarbeit darin, einen Mangel heutiger Geo-Informationssysteme zu beseitigen: sie bieten zwar umfangreiche Funktionalität zur Visualisierung raumbezogener Datensätze, enthalten aber nicht das erforderliche Wissen um diese Funktionalität sinnvoll zu nutzen. Dieser Mangel äußert sich in der wachsenden Zahl mit GIS erstellter Visualisierungen, die raumbezogene Datensätze ungenau oder sogar verfälscht darstellen und damit räumliche Planung behindern und falsche Entscheidungen hervorrufen können.

Um den als „Wissenslücke in der GIS-Visualisierung“ charakterisierten Mangel zu beheben, werden Elemente eines integrierten Benutzerunterstützungssystems untersucht, konzipiert und prototypisch realisiert. Der Schwerpunkt dieser Arbeit liegt auf der Automatisierung des GIS-Visualisierungsentwurfs unter Einsatz graphischen Wissens und unter Berücksichtigung der Zielsetzung und des Kontexts der Visualisierung. Das Benutzerunterstützungssystem soll in der Lage sein, ausdrucksfähige und effektive Visualisierungen von Datensätzen automatisch zu entwerfen, um Benutzern, die nicht über umfangreiches graphisches Vorwissen verfügen, eine möglichst fehlerfreie und effektive Darstellung zu garantieren.

Kapitel 2 dieser Arbeit führt in die **Problematik der Effektivität und Nutzbarkeit der GIS-Visualisierung** detailliert ein. Die beiden zentralen Begriffe dieser Arbeit, Geo-Informationssysteme und Visualisierung werden eingeführt und der aktuelle Stand der Technik vorgestellt. Auf die vielfältigen Möglichkeiten, die GIS-Visualisierung den Anwendern von Geo-Informationssystemen bietet, wird im Detail eingegangen, aber auch auf die Schwierigkeiten und Fehlerquellen, die diese komplexen Werkzeuge mit sich bringen. Das Kapitel schließt mit der Vision eines integrierten Systems zur Benutzerunterstützung, das Anwendern von GIS die Erstellung von nützlichen Visualisierungen von Geodaten garantiert.

Kapitel 3 gibt einen Überblick über wichtige Vorarbeiten und den **Stand der Technik** auf dem Gebiet **Automatisierter Visualisierungsentwurf**. Es werden einflußreiche Forschungsarbeiten auf diesem Gebiet vorgestellt und die Implikationen für diese Arbeit gezogen. Weitere Vorarbeiten auf dem Gebiet der Charakterisierung von Datensätzen und Visualisierungszielen werden diskutiert. Das dieser Arbeit ebenfalls verwandte Gebiet des automatisierten kartographischen Entwurfs wird vorgestellt und für den GIS-Visuali-

sierungsentwurf wichtige Ergebnisse festgehalten. Schließlich werden die für den automatisierten GIS-Visualisierungsentwurf bestehenden offenen Probleme vorgestellt.

Kapitel 4 untersucht, welche variablen Faktoren die Effektivität von GIS-Visualisierungen bestimmen und ordnet sie in ein **Modell zur Charakterisierung von Daten und Kontext der GIS-Visualisierung** ein. Neben der Charakterisierung von GIS-Sachdaten ermöglicht dieses Modell auch die Beschreibung der Zielsetzung der Visualisierung, der Ausgabemedien und der Fähigkeiten bzw. Vorkenntnisse des Benutzers. Das Modell stellt gleichzeitig einen Ausgangspunkt für die Automatisierung des Visualisierungsentwurfs dar.

Kapitel 5 stellt eine **Methodik für den automatisierten GIS-Visualisierungsentwurf** vor. Die Methodik ermöglicht den Entwurf ausdrucksfähiger Visualisierungen von GIS-Datensätzen unter Berücksichtigung der Aufgabenstellung und des Kontexts. Sie umfaßt drei Hauptbestandteile: (1) ein Algorithmus für den Entwurf ausdrucksfähiger Visualisierungen raumbezogener Datensätze, (2) ein Schema zur optimalen Vergabe visueller Ressourcen und (3) ein Fuzzy-Schema für die Effektivitätsabschätzung von Visualisierungen.

In Kapitel 6 wird die **empirische Bestimmung der Effektivität einiger GIS-Visualisierungsformen** diskutiert. In einer Reihe psychologischer Tests wurde die Effektivität verschiedener Visualisierungsformen für unterschiedliche Aufgaben und in einem Standardkontext bestimmt und damit eine Grundlage für die Fuzzy-Bewertung in praktischen Systemen geschaffen.

Kapitel 7 beschreibt **VIZARD, eine prototypische Realisierung der intelligenten Umgebung zur Benutzerunterstützung**. Die Ziele dieser Entwicklung und der Funktionsumfang des Systems werden vorgestellt und die Wissensbasis mit Regeln zur Ausdrucksfähigkeit und Effektivität von Visualisierungstechniken sowie die graphische Benutzerschnittstelle besprochen.

Kapitel 8 faßt die **Ergebnisse** der Arbeit zusammen, bewertet sie und zeigt, inwieweit sie sich verallgemeinern und auf verwandte Anwendungen bzw. Probleme übertragen lassen.

# Kapitel 2

## Die Problematik: Effektivität und Nutzbarkeit der GIS-Visualisierung

Kapitel 2 führt in die Problematik der Effektivität und Nutzbarkeit der GIS-Visualisierung detailliert ein. Zunächst werden die beiden zentralen Begriffe dieser Arbeit, Geo-Informationssysteme und Visualisierung definiert und jeweils der aktuelle Stand der Technik diskutiert. Die Darstellung geht nur bei denjenigen Aspekten von GIS und Visualisierung in die Tiefe, die für die vorliegende Arbeit relevant sind. Anschließend wird eine Synthese durchgeführt und ein neuer Begriff geprägt: *GIS-Visualisierung*. Auf die vielfältigen Möglichkeiten, die GIS-Visualisierung den Anwendern von Geo-Informationssystemen bietet, wird im Detail eingegangen, aber auch auf die Schwierigkeiten und Fehlerquellen, die deren komplexe Funktionalität mit sich bringt. Das Kapitel schließt mit der Vision einer Softwareumgebung, die Anwender von GIS bei der Erstellung nützlicher Visualisierungen von Geodaten unterstützt.

### 2.1 Geo-Informationssysteme: Stand der Technik

Geo-Informationssysteme dienen der Verwaltung und Analyse raumbezogener Information. Der Begriff Geographisches Informationssystem wurde bereits in den sechziger Jahren erstmals verwendet, um ein neuartiges, raumbezogenes Informationssystem zu bezeichnen [MaGR91]. Heute zählen GIS zu den komplexesten Informationssystemen, die auf dem Markt angeboten werden. Sie verfügen häufig über eine sehr umfassende und schwer überschaubare Funktionalität, so hat beispielsweise ein weit verbreitetes, kommerzielles GIS ca. 1800 unterschiedliche Funktionen [Fran93]. Die Komplexität dieser Systeme wird in folgender Definition deutlich, die alle wesentlichen Komponenten und Funktionalitäten aufzählt:

Ein Geo-Informationssystem ist ein rechnergestütztes System, das aus Hardware, Software, Daten und Anwendungen besteht. Mit ihm können raumbezogene Daten digital erfasst und registriert, gespeichert und reorganisiert, modelliert und analysiert sowie alphanumerisch und graphisch präsentiert werden. [BiFr91]

Im Laufe dieses Kapitels wird ein Überblick über wichtige Aspekte der Geo-Informationssysteme gegeben. Dazu wird der Aufbau der Softwarekomponente und der Daten in einem GIS erklärt. Die zentralen Anwendungsgebiete von GIS sowie einige aktuelle Themen der GIS-Forschung bilden den Abschluß dieses Kapitels.

### 2.1.1 Aufbau und Operationen eines GIS

#### Architektur von GIS

Die Module der GIS-Software bauen, wie in allen Informationssystemen, sehr stark auf einer Datenbankschnittstelle auf. Der modulare Aufbau eines typischen Geo-Informationssystems wurde schon in Abb. 1-1 gezeigt. Im Zentrum steht eine Geodatenbank, die raumbezogene Daten effektiv speichern kann [Fran91]. Auch die Datenzugriffsmethoden sind für Geodaten optimiert, d.h. ein effizientes Retrieval von Datensätzen nach Kriterien wie der absoluten, räumlichen Lage und der räumlichen Nachbarschaft ist gewährleistet. Eine spezielle Datenakquisitionskomponente ermöglicht die Eingabe und Aktualisierung des Geodatenbestandes, wobei häufig Schnittstellen zu besonderen Eingabegeräten wie Scannern und Digitalisiertischen unterstützt werden. Datensätze können im GIS mit besonderen Analyse- und Verarbeitungsfunktionen ausgewertet und kombiniert werden. Die Visualisierungs- und Präsentationskomponente stellt die Geodatensätze oder die Analyseergebnisse schließlich in graphischer und textueller Form dar. Alle diese Komponenten kann der GIS-Anwender von einer einheitlichen Benutzerschnittstelle aus steuern. Alle heute verwendeten GIS haben eine graphische Benutzerschnittstelle [Medy94, Turk90].

Elemente eines GIS sind neben der Software natürlich auch Hardware und Daten. Typische, GIS-spezifische Hardwarekomponenten sind großformatige Scanner und Plotter, mit denen sich Karten bis zum Format DIN A0 erzeugen und bearbeiten lassen. Für die Zwecke der Datenerfassung gibt es eine Reihe von vermessungstechnischen und photogrammetrischen Geräten. Neben den bereits erwähnten Digitalisiertischen sind auch Empfänger für das Satelliten-Navigationssystem GPS eine wichtige Datenquelle. Daten spielen in Geo-Informationssystemen eine besondere Rolle, da sie ein wesentlicher Kostenfaktor sind (bei manchen Anwendungen bis zu 80% der Gesamtprojektkosten) und die Ergebnisse eines GIS natürlich immer nur so präzise und vertrauenswürdig wie die zugrundeliegenden Daten sein können.

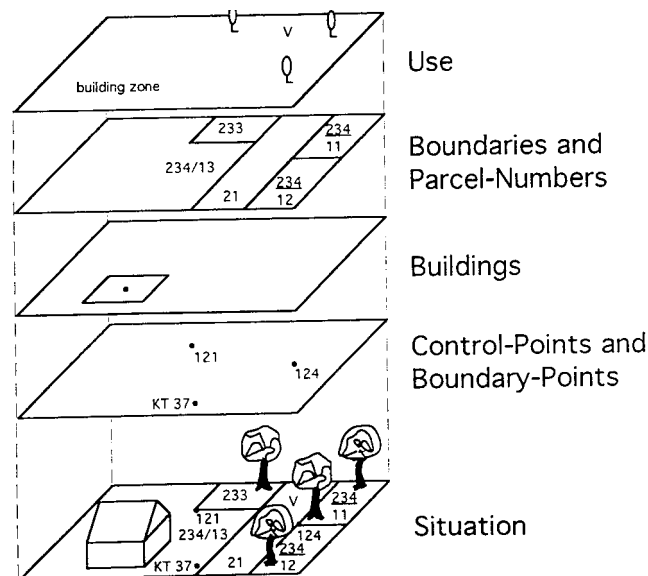
#### Karten und Geo-Informationssysteme

Geo-Informationssysteme ersetzen oder vereinfachen viele manuelle Arbeiten, die früher mühsam mit Papierkarten und am Zeichentisch durchgeführt wurden. Geodatenbanken haben die Karte als primären Geo-Informationsspeicher verdrängt und sind in der Lage, Geo-Information wesentlich detaillierter und exakter zu repräsentieren. Karten werden heute aus Geodatenbanken in einem semiautomatischen Prozeß erzeugt. Jede Karte stellt nur eine von mehreren möglichen Sichten auf die Geo-Information dar.

Außer der Rolle als Informationsspeicher und Visualisierungsform haben GIS von der Karte aber noch andere Aufgaben übernommen. Eine dieser klassischen Arbeitsformen mit Karten, die Überlagerung, taucht auch in modernen GIS in ähnlicher Form auf. Mit



einer Überlagerung lassen sich zwei verschiedene Sichten eines Ausschnitts der realen Welt - bzw. zwei Datensätze mit dem gleichem geographischen Bezugsraum - graphisch miteinander kombinieren. Das Ergebnis ist eine neue Sicht - bzw. ein neuer Datensatz - der die Beziehung der beiden zueinander dokumentiert. Solche Überlagerungen sind ein sehr effektives Hilfsmittel für raumbezogene Analysen und Planungen. Abb. 2-1 zeigt ein Beispiel dafür, wie aus der Überlagerung verschiedener thematischer Sichten eine neue, spezifische Sicht entsteht.



**Abb. 2-1:** Kartographische Überlagerung von Layern [Fran93]

### GIS-spezifische Operationen

Operationen von GIS sind für die Verarbeitung raumbezogener Daten optimiert. Ein typisches *general purpose* GIS stellt eine vergleichsweise hohe Zahl eigenständiger Operationen zur Verfügung (1800 Operationen wurden schon als Beispiel genannt). Die hohe Zahl kommt auch dadurch zustande, daß GIS mit sehr unterschiedlichen Repräsentationen von Geo-Information arbeiten müssen, wie z.B. Objekt-orientierten und Raster-orientierten Repräsentationen.

Die GIS-Operationen lassen sich, entsprechend der Architektur der GIS-Software, vier Aufgabenbereichen zuordnen:

- *Erfassung und Modellierung:* Erstellung eines anwendungsspezifischen Geodatenschemas, Aufnahme von Vermessungs- und Erfassungsergebnissen in die Geodatenbank und interaktive Erstellung und Modifikation der Geodaten.
- *Abfrage:* Formulierung von Queries auf der Geodatenbank nach räumlichen Kriterien wie geographischer Lage, geographischer Nachbarschaft und topologischer Verwandtschaft.

- *Verarbeitung und Analyse:* Kombination, Verschneidung und Überlagerung von Geodaten mit unterschiedlicher Repräsentation und unterschiedlichem Raumbezug, statistische Auswertung quantitativer, raumbezogener Daten und Datenauswahl nach komplexen Kriterien.
- *Präsentation und Visualisierung:* Darstellung von räumlichen Datensätzen und Analyseergebnissen in Form von Texten, Listen, topographischen und thematischen Karten, Diagrammen und Bildern.

## 2.1.2 GIS-Daten und -Datenmodelle

### Geodaten

Geodaten sind digitale Repräsentationen eines Ausschnitts der realen Welt. Sie unterscheiden sich von Modelldaten aus CAD-Systemen durch den Maßstab bzw. durch die Größe des repräsentierten Ausschnitts und auch durch ihre Dimension. Geodaten haben typischerweise einen Maßstab zwischen 1:1 000 und 1:1 000 000 und sind zweidimensional. Dreidimensionale Geodaten gewinnen jedoch stetig an Bedeutung [CoJu98].

Repräsentiert werden materielle Gegenstände der Welt, aber auch gedachte Gegenstände wie politische Grenzen bzw. Verwaltungsgrenzen. Im großmaßstäblichen Bereich beschreiben die Geodaten Gebäude, Verkehrswege, Flüsse, Gemeindegrenzen, usw., während im kleinmaßstäblichen Bereich Städte, Landkreise oder Verwaltungsbezirke die Bezugsgrößen sind.

Zwei prinzipielle Repräsentationsformen für Geodaten werden unterschieden:

- *Vektordaten:* Sie beschreiben die Ausdehnung räumlicher Gegenstände. Drei Klassen von Gegenständen werden unterschieden: punktförmige (z.B. Verkehrszeichen), linienförmige (z.B. Straßen) und flächenförmige (z.B. Gebäude). Je nach Maßstab und Anwendungen können die Gegenstände natürlich auch anders dargestellt werden, z.B. ein Gebäude im kleinen Maßstab auch als Punkt. Punktförmige Gegenstände werden durch die Koordinaten eines Punktes repräsentiert. Linienförmige und flächenförmige Gegenstände können beispielsweise durch Punktfolgen repräsentiert werden, wobei die Punkte bei linienförmigen Objekten zu einer Kurve verbunden werden und bei flächenförmigen diese Kurve zusätzlich geschlossen wird.
- *Rasterdaten:* Bei dieser Repräsentationsform wird ein geographischer Bereich ausschließlich durch Flächen modelliert. Dazu wird er mit einem regulären Gitter in quadratische Zellen unterteilt und jeder Zelle genau ein Attributwert zugeordnet. Zur primären Datenerfassung werden Rasterdaten vor allem im klein- und mittelmaßstäblichen Bereich eingesetzt. Rasterdaten sind in der Regel über Luft- oder Satellitenaufnahmen erfaßt.

Vektor- und Rasterdaten haben beide ganz spezifische Vor- und Nachteile. Der entscheidende Vorteil der Vektordaten ist der Objektbezug: Vektordaten geben Auskunft über Ausdehnung und räumliche Lage der Gegenstände in der Welt. Rasterdaten dagegen ent-



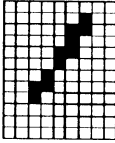

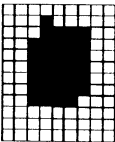
Element	Vektor		Raster	
	Digital	Analog	Digital	Analog
Punkt	x,y Koord.	•	Pixel	
Linie	x y Koord.-folge		Pixel	
Fläche	geschlossene x y Koord.-folge		Pixel	

Abb. 2-2: Vektor- und Rasterdaten [BiFr91]

halten keine explizite Information über Geometrien oder Ausdehnungen. Vorteilhaft an den Rasterdaten ist, daß sie sich sehr einfach und kostengünstig erfassen lassen und flächenhafte, natürliche Phänomene, wie z.B. Vegetation, besser repräsentieren. Dagegen ist die Erfassung von Vektordaten äußerst kostenintensiv und mit ihnen lassen sich Bauwerke, Verkehrsnetze, Grenzen, etc. besser repräsentieren. Beide Datenarten können in einem GIS gemeinsam verwendet werden, aber insgesamt sind die Vektordaten wichtiger, da sie die Eigenschaft des Objektbezugs mit sich bringen. Rasterdaten dienen auch häufig als Grundlage für die Erzeugung von Vektordaten.

Stattet man Vektordaten, die die Geometrie räumlicher Gegenstände wiedergeben, mit zusätzlicher Information in Form von Attributen und topologischen Verknüpfungen aus, dann entstehen *Geo-Objektdaten*. Diese beschreiben den räumlichen Gegenstand aus der Sicht einer Anwendung vollständig. Um einen Gegenstand „Straße“ vollständig zu repräsentieren, müßten beispielsweise zu der Vektorinformation über die geographische Lage und Ausdehnung noch Attributinformationen mit den Namen der Straße sowie topologische Verknüpfungen zu allen Gebäuden in der Straße aufgenommen werden.

### Raumbezogene Sachdaten

Eine weitere Klasse von Daten im GIS sind die raumbezogenen Sachdaten. Wie ihr Name vermuten läßt, beziehen sie sich auf räumliche Objekte, geben in der Regel aber keine Positionen und Ausdehnungen an. Statt dessen geben sie z.B. weitere Attribute eines Gegenstandes an, der schon als Geo-Objekt modelliert ist. Sie *beziehen* sich also auf Geo-Objekte. Sind beispielsweise die Straßen einer Stadt als Geo-Objekte modelliert, kann das

Verkehrsaufkommen als Sachdatensatz bezüglich dieser Geodaten repräsentiert werden. Jedes Element dieses Sachdatensatzes ist dann ein Wertepaar, bestehend aus einer Anzahl gezählter Fahrzeuge und einer Referenz auf eine Straße, beispielsweise über den Straßennamen.

Raumbezogene Sachdaten benötigen also immer einen Geo-Objektdatensatz, auf den sie sich beziehen. Die Sachdaten selbst können beschreibend sein, wie z.B. eine Grundbuchdatei mit Angaben über die Eigentümer von Flurstücken. Sachdaten können aber auch quantitativ sein und numerische Größen, wie das Verkehrsaufkommen im obigen Beispiel, repräsentieren. Quantitative Sachdaten spielen eine zentrale Rolle in vielen GIS-Anwendungen in der räumlichen Analyse und Planung.

### 2.1.3 Wichtige GIS-Anwendungsgebiete

Dieser Abschnitt stellt die drei Hauptanwendungsgebiete von GIS kurz vor: Planung, Forschung und Verwaltung. Für Anwendungen in der Forschung wird zusätzlich ein konkretes Beispielprojekt vorgestellt und die besondere Rolle von GIS aufgezeigt. Zum Abschluß wird ein Ausblick auf die zukünftige Entwicklung der GIS und ihrer Anwendungen gemacht.

#### **Planung**

GIS unterstützen heute alle Phasen einer räumlichen Planung. Ein GIS-Projekt im Planungsbereich beginnt mit der Zusammenstellung der Daten über das Planungsgebiet. Als Basis der GIS-Datenbank müssen zunächst Geo-Objektdaten des Gebietes in geeignetem Maßstab und ausreichender Präzision beschafft werden. Immer häufiger sind solche Daten schon am Geodatenmarkt verfügbar. Basierend auf diesen Geo-Objektdaten werden anschließend die benötigten Sachdaten erfaßt. Bei der Planung einer neuen Straße werden beispielsweise Sachdaten über das Verkehrsaufkommen und die Lärmschutzgebiete benötigt. Mit Hilfe des GIS wird im nächsten Planungsschritt die Situation analysiert und verschiedene Lösungsvorschläge (z.B. Straßenverläufe) hergeleitet und dokumentiert. Das GIS hilft hier auch bei der Bewertung der Alternativen und bei der graphischen Präsentation des Planungsergebnisses.

Eine Aufzählung aller Anwendungsbereiche, in denen GIS heute erfolgreich für Planungsaufgaben eingesetzt werden, enthält unter anderem:

- Städteplanung, z.B. Trassen von Verkehrswegen
- Regionalplanung, z.B. gezielte Infrastrukturmaßnahmen
- Umweltplanung, z.B. in der Auswertung von Satellitenaufnahmen
- Absatzplanung, z.B. die Suche nach geeigneten Standorten für Filialen
- Routenplanung, z.B. für Lieferanten
- Einsatzplanung, z.B. von Feuerwehr und Rettungswagen

**Forschung**

GIS werden in vielen Disziplinen für Forschungsarbeiten mit räumlichen Bezug eingesetzt. Beispiele solcher raumbezogener Forschung mit GIS-Unterstützung sind die Untersuchung von Migrationsbewegungen in der Humangeographie, die Seuchenforschung in der Medizin und die Wahlforschung in der Statistik.

Bei Forschungsanwendungen steht die Analyse des Studiengebietes und die graphische Präsentation im Vordergrund. GIS helfen durch die explorative Analyse raumbezogener Datensätze, Hypothesen über Entwicklungen und Zusammenhänge räumlicher Prozesse zu formulieren. Im nächsten Schritt unterstützen sie den Forscher beim Bestätigen oder Widerlegen von Hypothesen. Am Schluß steht die graphische Präsentation von Forschungsergebnissen in Form von thematischen Karten und Diagrammen. Moderne GIS produzieren hochwertige Graphiken, die für Berichte, Publikationen und Präsentationen gleichermaßen eingesetzt werden.

Ein Beispiel für die Anwendung von GIS in raumbezogener Forschung ist ein interdisziplinäres Projekt, das vom Institut für Geschichte und dem Fachgebiet Graphisch-Interaktive Systeme der Technischen Hochschule Darmstadt gemeinsam durchgeführt wurde [ScJu94, Jung94]. In diesem Forschungsprojekt wurde das Historische Stadtinformationssystem (HIST), ein fachspezifisches GIS für die raumbezogene historische Sozial- und Stadtentwicklungsforschung entwickelt und angewendet. Eingesetzt wurde HIST im DFG-Forschungsprojekt „Elektrifizierung und Stadtentwicklung in südwestdeutschen Städten 1880-1935“ [BöSc93], das den Prozeß der Elektrifizierung der Stadt als zentrale Infrastrukturinnovation in seinen Erscheinungsformen und Auswirkungen auf die Stadtentwicklung untersuche.

HIST resultierte aus den räumlichen Dimensionen des Forschungsgegenstandes. Im Zuge von Stadtwachstum wie auch dem Einsatz technischer Infrastruktursysteme vollzog sich um die Jahrhundertwende in deutschen Städten ein raumstruktureller Wandel. Bebauung und Bodenwerte veränderten sich, Nahverkehrssysteme schafften neue Zonen von Erreichbarkeiten, die netzförmige Versorgung mit Wasser und Energie ermöglichte neue Standorte für Industrie und Gewerbe. Mit HIST wurde die die Stadtentwicklung als Wandel der Raumstruktur analysiert, insbesondere hinsichtlich der räumlichen Wirkungen der technischen Infrastrukturen.

HIST ermöglicht die Integration und Visualisierung räumlicher Daten mit Daten zur Sozial-, Wirtschafts- und Bevölkerungsstruktur auf verschiedenen Aggregationsebenen. Dadurch wird vor allem ein höherer Grad der Anschaulichkeit von Raumstrukturen und deren Wandel im Zuge der Urbanisierung ermöglicht. Über die kartographische Präsentation von Forschungsergebnissen hinaus bietet das System auch die Möglichkeit, Hypothesen über den Zusammenhang von Variablen „durchzuspielen“.

**Öffentliche Verwaltung**

Die frühesten Anwendungen von GIS lagen im Bereich der öffentlichen Verwaltung. Die Erfassung und Verwaltung von Geodaten und raumbezogener Information ist besonders in Deutschland auch heute noch ein äußerst aktives Anwendungsfeld von GIS. Durch den



Durch die globalen Datennetze und das zunehmende Vordringen der Informations- und Kommunikationstechnik in den privaten Bereich, wird die Anwendungspalette von GIS aller Wahrscheinlichkeit nach noch zunehmen. Jean-Claude Muller gab 1993 die Schätzung ab, im Jahr 2000 würden mindestens eine Million Anwender regelmäßig GIS benutzen [Mull93], bei anhaltendem Ausbau der Informationsinfrastruktur rechnet er mit einer wesentlich größeren Zahl.

Drei neue Arten der GIS-Anwendung zeichnen sich schon heute ab und werden in Zukunft ein beträchtliches Potential für den GIS-Markt darstellen:

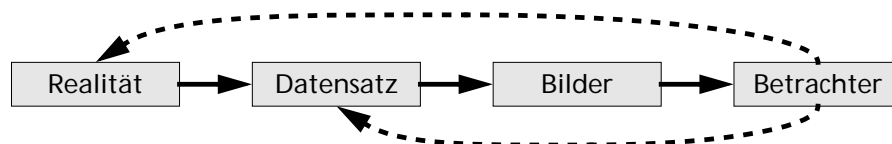
- *Navigationssysteme:* Sie werden in Fahrzeugen eingesetzt und erleichtern Fahrern die Orientierung durch gezielte Hinweise auf Wege, Umleitungen und touristisch interessante Punkte. Insbesondere zusammen mit GPS-Empfängern ermöglichen sie eine präzise und flexible Navigation.
- *Points-of-information:* Diese Informationssäulen werden zunehmend auf Bahnhöfen, Flughäfen, in Behörden, etc. postiert. Sie dienen als Quelle für aktuelle, raumbezogene Information über Verkehrsverbindungen, Wege und Standorte.
- *WWW-basierte GIS:* Das Internet und das World-Wide-Web sind das Medium dieser sich rapide entwickelnden GIS-Anwendungen. Raumbezogene Information kann auf diesem Weg aktuell und global verfügbar gemacht werden. Umweltdaten lassen sich beispielsweise schon heute von einigen Institutionen auf diesem Weg abrufen.

## 2.2 Computervisualisierung: eine neue wissenschaftliche Fachrichtung

Der neue Begriff „Computervisualisierung“ [Domi96] hat sich noch nicht etabliert, eignet sich aber sehr gut als Obergriff für eine Reihe von Spielarten der Visualisierung, die sich zwar jeweils durch spezifische Techniken und Regeln voneinander unterscheiden, insgesamt aber einer einheitlichen Zielsetzung und Methodik unterworfen sind. Die Zielsetzung, digitale Information mit Computerprogrammen in Bilder umzuwandeln und die Methodik des graphischen Entwurfs charakterisieren die Computervisualisierung als neue wissenschaftliche Fachrichtung.

Computervisualisierung wandelt Datensätze in Bilder um, mit dem Ziel, das Verständnis dieser Datensätze zu erleichtern. Sie kann Phänomene sichtbar machen, die in der ursprünglichen Repräsentation unsichtbar waren. Sie nutzt dabei die Eigenschaft aus, daß das menschliche visuelle System viele Operationen, die für die Interpretation von Datensätzen hilfreich sind, an Bildern sehr effektiv ausführen kann. So wird beispielsweise in Börsenberichten die Entwicklung eines Aktienindex immer in Form eines Liniendiagramms dargestellt, da sich an ihm Trends und Schwankungen viel einfacher erkennen lassen als in einer Wertetabelle.

Das Anliegen der Computervisualisierung läßt sich mit einem einfachen Diagramm darstellen (siehe Abb. 2-4). Aus der Realität werden einige Phänomene ausgewählt und in einem Computer durch einen digitalen Datensatz repräsentiert. Die Datensätze können beispielsweise durch Messungen oder Simulationen gewonnen werden. Sie werden mit einem Visualisierungsverfahren in Bilder transformiert und dargestellt. Zwei Formen der Rückkopplung sind in diesem Prozeß hervorzuheben. Zunächst lassen sich Aussagen über die Charakteristika des dargestellten Datensatzes machen. Der Anwender der Visualisierung kann aber durch die Interpretation der Darstellung auch Rückschlüsse auf die Realität ziehen. Ein erfahrener Börsenmakler erkennt in der Entwicklung des Aktienindex charakteristische Trends und Schwankungen. Er kann daraus in einem weiteren Interpretationsschritt künftige Kursentwicklungen prognostizieren.



**Abb. 2-4:** Das Anliegen der Computervisualisierung

In diesem Visualisierungsprozeß spielt natürlich auch die Computergraphik eine große Rolle. Diese Fachrichtung beschäftigt sich mit den Techniken der Visualisierung und der Darstellung von Bildern und Simulationen, wobei die Korrektheit und die Zeit-Effizienz der Techniken im Vordergrund stehen. Im Gegensatz dazu beschäftigt sich die Fachrichtung Computervisualisierung mit der Frage, welche Visualisierungstechniken unter den gegebenen Bedingungen für einen Datensatz geeignet sind. Bei der Computervisualisierung stehen also die visuelle bzw. kognitive Effektivität der Techniken im Vordergrund [Domi96, FJSD96].

### 2.2.1 Formen und Anwendungsgebiete der Computervisualisierung

Wie bereits dargelegt wurde, ist Computervisualisierung ein Oberbegriff für eine Reihe von Visualisierungsformen, die auf eine unterschiedlich lange Tradition zurückgehen. Zu den wichtigsten Formen der Computervisualisierung zählen:

- Die *technisch-wissenschaftliche Visualisierung* (Visualization in Scientific Computing, ViSC), entstanden Ende der achtziger Jahre, als immer größere Datenmengen aus physikalischen Messungen und Simulationen anfielen, die mit klassischen Methoden nicht mehr bewältigt und interpretiert werden konnten. Charakteristisch für die technisch-wissenschaftliche Visualisierung ist der intensive Einsatz dreidimensionaler Methoden und die dafür benötigten Hochleistungs-Graphikrechner.



- Die *Informationsvisualisierung* (information visualization), entstanden in den neunziger Jahren als Reaktion auf die ebenfalls kaum mehr überschaubare Informationsflut in vernetzten Datenbanken und im Internet. Sie erzeugt dynamische visuelle Darstellungen von Datenbankstrukturen, die das Auffinden gesuchter Information erleichtern.
- Die *Computergestützte Kartographie* (computer aided cartography), entstanden schon in den Siebzigern und auch heute noch die vorherrschende Visualisierungsform in GIS.

Für die Anwendungsfelder und -formen der Computervisualisierung soll hier ein konkretes Forschungsprojekt als Beispiel dienen. An der Hydrographischen Abteilung des Instituto Superior Tecnico in Lissabon werden Methoden der numerischen Simulation von Meeresströmungen in den Mündungsbereichen von Flüssen entwickelt. Die zwei- und dreidimensionalen numerischen Modelle dienen zur Vorhersage der Auswirkungen von Infrastrukturmaßnahmen (z. B. Brückenneubaten) auf die Ökosysteme Fluß und Meer. Die Simulationen müssen ständig angepaßt und verbessert werden, damit sie ein realistisches Modell der realen Vorgänge abgeben [Neve88, Jung93]. Um die Interpretation der Simulationsergebnisse, insbesondere bei 3D-Modellen zu vereinfachen, setzen die Forscher Techniken der technisch-wissenschaftlichen Visualisierung ein. In Abb. 2-5 wurde zum Beispiel das Verhalten von Partikeln in einer Strömung durch eine Schleuse simuliert, und die Partikelbahnen verfolgt. Mit solchen Visualisierungen läßt sich untersuchen, wie realistisch ein numerisches Modell natürliche Strömungsvorgänge repräsentiert.



Abb. 2-5: Visualisierung eines 3D-Strömungsfeldes mittels Partikelbahnen [Jung93]

### 2.2.2 Wissensquellen der Computervisualisierung

Zielsetzung der Computervisualisierung ist vor allem, eine visuelle Repräsentation eines Datensatzes zu finden, die den Datensatz korrekt darstellt und seine Interpretation erleichtert. Damit dieses Ziel erreicht werden kann, müssen mehrere Voraussetzungen geschaffen werden. Zum einen muß man sich klar machen, welche Eigenschaften des Datensatzes interpretiert werden und welche der visuellen Interpretationsformen im Vordergrund stehen sollen. Zusätzlich muß man wissen, welche Visualisierungstechniken bestimmte Eigenschaften eines Datensatzes ausdrücken und wie ihre Interpretation durch das menschliche visuelle System ermöglicht werden kann.

Um dieses Ziel zu erreichen, setzt die Computervisualisierung Wissen aus verschiedenen anderen Fachgebieten ein:

- aus der *Kartographie* die graphische Semiotik, also die Lehre von der Bedeutung und der Grammatik graphischer Zeichen
- aus der *mathematischen Statistik* Kenntnisse über Darstellungsformen der numerischen und statistischen Eigenschaften von Datensätzen
- aus der *Psychologie* Erkenntnisse über die menschliche visuelle Wahrnehmung
- aus *Kunst, Design und Architektur* die ästhetischen und funktionalen Prinzipien des guten graphischen Entwurfs
- aus den verschiedenen *Anwendungsbereichen* Wissen über spezifische Konventionen, Techniken und Erfordernisse der Darstellung

## 2.3 Der Begriff „GIS-Visualisierung“

Mit der Einführung der computergestützten Kartographie und der GIS ab dem Ende der siebziger Jahre, fand auch eine beträchtliche Erweiterung der kartographischen Techniken statt. Viele der Darstellungsmethoden, die bisher als aufwendig galten und deswegen wenig Verwendung fanden, waren mit Computerunterstützung plötzlich leicht zu realisieren [SIEg91]. Um diese Zeit traten beispielsweise unklassifizierte Choroplethenkarten erstmals häufiger auf. Die für sie charakteristischen glatten Farbverläufe konnten mit Zeichenstiften oder Schablonen vorher nicht realisiert werden, waren mit dem Computer aber einfach herzustellen. Ebenso waren jetzt sogenannte Diagrammkarikaturen (Engl.: cartograms), die durch die Verzerrung geographischer Flächen eine Größe wie z.B. die Bevölkerungszahl der Länder ausdrücken, realisierbar und wurden häufiger verwendet. Durch numerische Berechnungsverfahren, wie die nichtlineare zweidimensionale Regression, wurden auch die Möglichkeiten der Datenklassifizierung wesentlich erweitert [BuMa91b].

Ab dem Ende der achtziger Jahre hielten dann dreidimensionale Darstellungen [Kraa90] und die Techniken der technisch-wissenschaftlichen Visualisierung ihren Einzug in die Geo-Informationssysteme [Weib94, FIHE96, ScUr97]. Gründe dafür waren einerseits die

enormen Datenmengen, die durch moderne Fernerkundungsmethoden und Satellitenbeobachtung gewonnen werden konnten. Neuartige Anwendungen, insbesondere im Umweltbereich und der Simulation, arbeiteten mit Vektorfeldern und 3D-Daten und erforderten deswegen spezielle Visualisierungsformen. Diese Entwicklung wurde durch die breite Verfügbarkeit leistungsfähiger Graphikworkstations noch verstärkt. Betrachtet man die Palette der Visualisierungstechniken, die in Umwelt-Informationssystemen heute eingesetzt werden [Denz93], findet man eine „bunte“ Mischung klassischer thematischer Karten mit Visualisierungs- und Simulationsmethoden.

Ende der Neunziger erweitern zwei neue Strömungen die Techniken und Anwendungsgebiete der GIS-Visualisierung. Einerseits bringt die Entwicklung der Virtuellen Realität neue Dimensionen von Interaktivität in alle graphischen Anwendungen, also auch in die Computervisualisierung und in die GIS [CoJu98]. Andererseits steht mit der Entwicklung der globalen Informations-Infrastruktur eine zweite Computerrevolution an: sehr viele Menschen werden Zugang zu raumbezogenen Informationen erlangen und dieses Potential auch nutzen wollen. Hier werden thematische Karten weiterhin eine große Rolle als anschauliche und vertraute Darstellungsform behalten. Sie müssen aber auf das neue Medium angepaßt werden, denn Bildschirmdarstellungen haben in ihrer Dynamik andere Charakteristika als beispielsweise Druckerzeugnisse.

Insgesamt läßt sich sagen, daß die Computerkartographie bereits zum Teil von einer umfassenderen „GIS-Visualisierung“ abgelöst worden ist. Mit der dargestellten Entwicklung wird sich dieser Trend sicher noch fortsetzen, was sich auch in der steigenden Anzahl wissenschaftlicher Publikationen zu diesem Thema zeigt. Im Jahr 1994 erschienen gleich zwei Kompendien zum Thema GIS-Visualisierung [HeUn94, MaTa94] und die International Cartographic Association (ICA) hat eine Kommission für Visualisierung gegründet [ICA].

Der Begriff GIS-Visualisierung hat sich bisher in der GIS-Forschung noch nicht etabliert, eignet sich aber gut als Oberbegriff dieser Methoden der Darstellung raumbasierter Information in GIS. Teilweise wird auch der Begriff „Kartographische Visualisierung“ verwendet, der aber weniger gut geeignet scheint, da er einen expliziten Schwerpunkt auf kartographische Methoden setzt.

Unter dem Begriff GIS-Visualisierung wird in dieser Arbeit die visuelle Darstellung raumbezogener Information mit den Methoden eines Geo-Informationssystems verstanden. Ziel der GIS-Visualisierung ist es, das Verständnis, die Dokumentation und die Kommunikation von räumlichen Zusammenhängen und räumlichen Prozessen zu ermöglichen. Zu den Techniken der GIS-Visualisierung zählen insbesondere thematische Karten, statistische Diagramme, dreidimensionale Darstellungen und die Animation.

## **2.4 Techniken der GIS-Visualisierung**

Dieses Kapitel gibt einen Überblick über die wichtigsten Techniken der GIS-Visualisierung. Der Schwerpunkt liegt bei den Methoden zur Darstellung quantitativer, raumbezogener Sachdaten.

### 2.4.1 Thematisch-kartographische Grundbegriffe

In der Kartographie werden zwei Klassen von Karten unterschieden: die topographischen und die thematischen Karten. Die *topographischen Karten* zeigen die Form, Gliederung und Bedeckung des Geländes. *Thematische Karten* hingegen zeigen nicht-topographische Erscheinungen, die jedoch einen geo-räumlichen Bezug haben, wie z.B. Bevölkerungsdichte, Verkehrsaufkommen, Landnutzung, etc. Eine wichtige Sondergruppe der thematischen Karten bilden die *statistischen Karten*, die im wesentlichen die Ergebnisse statistischer Zählungen oder Messungen mit ihrem räumlichen Bezug wiedergeben. [Imho71]

Alle Klassen von Karten, insbesondere aber die thematischen Karten, verwenden sogenannte *Signaturen* (engl.: *symbols*) als graphisches Grundelement. Signaturen sind genormte, abstrahierte Objektbilder oder Zeichen, die man graphischen Variationen (z.B. von Form, Farbe und Orientierung) unterziehen kann. Mit Signaturen sind stets auch gewisse Aussagen verbunden [Hake82]:

1. Immer eine qualitative Aussage (z.B. über die Art der Bodenbedeckung) durch die graphische Variation der Signatur,
2. meist auch eine Lageangabe (z.B. Orte und Grenzen),
3. oft auch eine quantitative Aussage (z.B. eine Bevölkerungsdichte), ebenfalls durch graphische Variation der Signatur.

Je nach der Dimension des räumlichen Bezugs unterscheidet man zwischen punktförmigen, linienförmigen und flächenförmigen Signaturen. Quantitäten können entweder durch gestufte oder durch stetige Signaturen wiedergegeben werden (siehe Abb. 2-6).

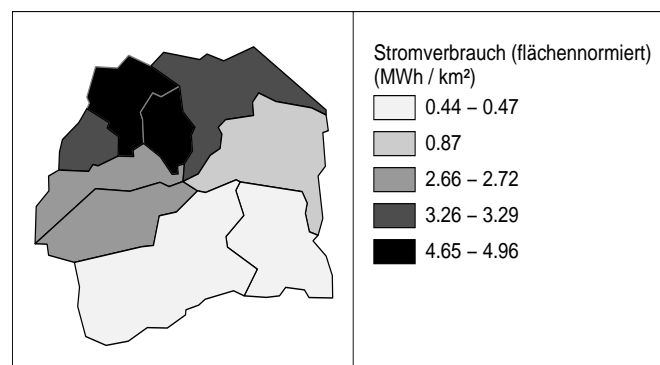
	punktförmig	linienförmig	flächenförmig
einfache Signaturen			
stetige Signatur (Quantitätsangabe)			
gestufte Signatur (Quantitätsangabe)			

Abb. 2-6: Signaturarten in der thematischen Kartographie

### 2.4.2 Choroplethen

Choroplethenkarten (engl. choropleth maps) sind eine der am häufigsten verwendeten Visualisierungstechniken aus der thematischen Kartographie. Sie werden von der Mehrzahl der Kartenleser gut verstanden und sind mit relativ einfachen Mitteln, auch ohne Computerunterstützung, herzustellen. Mit Choroplethen lassen sich Datensätze jedoch auch sehr verfälscht wiedergeben, sei es aus Unkenntnis der Datenklassifikationsmethoden, sei es um absichtlich ein verzerrtes Bild zu vermitteln.

Choroplethenkarten bilden eine quantitative Variable auf eine gestufte Flächensignatur ab. Am bekanntesten sind solche gestuften Signaturen, die die Intensität einer Farbe oder, wie in Abb. 2-7, die Helligkeit variieren. Voraussetzung für die Anwendbarkeit der Choroplethentechnik ist, daß die quantitative Variable nicht kontinuierlich über den geographischen Raum verteilt ist. Statt dessen muß der Raum in klar begrenzte Flächen aufgeteilt und jeder dieser Flächen ein diskreter, quantitativer Wert zugeordnet sein. Dies trifft zum Beispiel für die Aufteilung eines Gebietes in Bezirke zu. Wird jedem Bezirk ein quantitativer Wert wie der Stromverbrauch pro Flächeneinheit zugeordnet, läßt sich eine Choroplethenkarte des Stromverbrauchs erstellen (siehe Abb. 2-7).

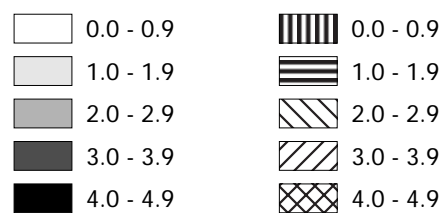


**Abb. 2-7:** Beispiel einer Choroplethenkarte

Eine wichtige Operation beim Entwerfen einer Choroplethenkarte ist die Klassifizierung der Daten. Der quantitative Datensatz wird nicht einer stetigen Signatur, wie einem Hell-Dunkel-Verlauf zugeordnet, sondern es werden Klassen von Datenwerten gebildet, denen jeweils genau eine Flächensignatur zugeordnet ist. In Abb. 2-7 ist dieser Zusammenhang in der rechten Spalte zu sehen: es sind fünf Klassen von Daten mit den fünf zugehörigen Flächensignaturen definiert. Daß bei Choroplethenkarten, wie auch bei anderen quantitativen Visualisierungstechniken, Daten klassifiziert werden, hat zwei wesentliche Gründe. Zum einen konnten vor dem Computerzeitalter stetige Signaturen kaum produziert und gedruckt werden. Zum anderen hilft die Klassifikation dem Kartenautor dabei, Aspekte der Datenverteilung durch eine bestimmte Klassenaufteilung zu betonen. Heute werden auch unklassifizierte Choroplethenkarten eingesetzt, vor allem in der Explorations- und Analysephase, wo eine möglichst originalgetreue Darstellung des Datensatzes gefordert ist.

Die Datenklassifizierung, also die Aufteilung des Wertebereichs des Datensatzes in diskrete Klassen, ist selbst eine der Hauptschwierigkeiten der Choroplethentechnik. Durch ungeschickte Aufteilung der Klassen kann der Eindruck eines Datensatzes verzerrt, im Extremfall sogar vollkommen verfälscht werden. Die Klassifizierung muß deswegen immer auf eine plausible methodische Grundlage gestellt werden [Monm91]. Eine Reihe numerisch-statistischer Verfahren helfen bei der Ermittlung geeigneter Datenklassifizierungen, ein guter Überblick über die Verfahren und ihre jeweilige Eignung findet sich bei Dent [Dent96, S. 129-138].

Auch bei der Zuordnung von Flächensignaturen zu den verschiedenen Datenklassen müssen einige Regeln berücksichtigt werden. Insbesondere muß die Ordnung der zugewiesenen gestuften Signatur der Ordnung der Klassen entsprechen. Da bei Choroplethen immer quantitative, und damit auch vollständig geordnete Daten abgebildet werden, muß auch die Signatur eine Ordnung ergeben. Abb. 2-8 zeigt zur Verdeutlichung zwei gestufte Signaturen, von denen die erste offensichtlich geeigneter ist als die zweite. Das liegt vor allem daran, daß die Ordnung der ersten Signatur der Ordnung der Klassifikation entspricht. Bei der zweiten Signatur besteht ein solcher Zusammenhang nicht (vgl. auch Abb. 1-2).

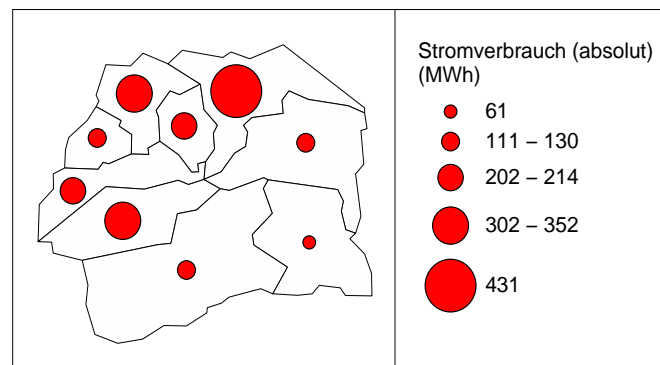


**Abb. 2-8:** Zwei gestufte Signaturen für Choroplethenkarten

### 2.4.3 Kreissignaturen

Kreissignaturenkarten (engl. proportional symbol maps) sind eine ebenfalls sehr häufig verwendete thematisch-kartographische Technik für quantitative Daten. Auch sie werden von Kartenlesern gut verstanden und ihre Produktion ist, ohne Computerunterstützung, sogar noch einfacher als die der Choroplethen. Wichtig ist bei den Kreissignaturenkarten die Wahl einer geeigneten Zuordnung von Signaturgrößen zu Datenklassen.

Kreissignaturenkarten bilden eine quantitative Variable auf eine gestufte punktförmige Signatur ab, die in der Größe variiert wird. (Neben Kreisscheiben werden gelegentlich auch andere punktförmige Signaturen wie Quadrate und Dreiecke verwendet.) Die quantitative Variable soll sich dabei auf isolierte Punkte oder klar begrenzte Gebiete beziehen. So beziehen sich z.B. Einwohnerzahlen von Städten ab einem gewissen Maßstab auf Punkte und können mit dieser Methode sehr gut dargestellt werden. Bezieht sich die quantitative Variable auf Gebiete, wird die Signatur z.B. jeweils an den Gebietsmittelpunkt gestellt.



**Abb. 2-9:** Beispiel einer Kreissignaturenkarte

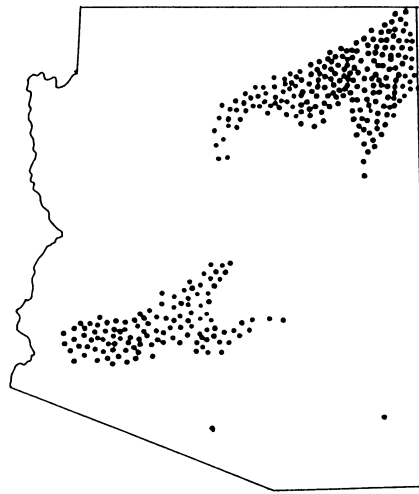
Die quantitative Variable wird in Klassen aufgeteilt und jeder Klasse eine Signaturgröße zugeteilt. Für die Klassenaufteilung gelten ganz ähnliche Regeln wie bei den Choroplethenkarten. Bei der Zuordnung der Signaturgrößen muß beachtet werden, daß die Größen gut voneinander unterschieden werden können, auch wenn sie an voneinander entfernten Stellen in der Karte stehen. Ein gewisser Mindestabstand zwischen den Signaturgrößen ist daher erforderlich. Die Größe der Signatur sollte natürlich mit der Größe des abgebildeten Datenwertes wachsen. Es muß aber darauf geachtet werden, daß die tatsächliche Fläche der Signatur proportional zu dem Wert ist, und nicht der Radius bzw. die Seitenlänge. Manche Kartographen argumentieren auch, daß einer Täuschung der menschlichen Wahrnehmung entsprechend kein linearer Zusammenhang bestehen sollte. Statt dessen soll die Quadratwurzel der Signaturfläche proportional wachsen. Es gibt hierüber in der kartographischen Literatur jedoch keine einheitliche Meinung [Dent96].

## 2.4.4 Weitere Formen thematischer Karten

### Punktkarten

Punktkarten (engl. dot maps) sind eine Form der Darstellung quantitativer Phänomene, die nicht an Punkten oder Flächen aggregiert sind, aber auch nicht im ganzen geographischen Raum auftreten (Beispiele solcher Phänomene sind Landnutzungsarten). Die Stärke der Punktkarten liegt darin, daß sie einerseits die Dichte eines Phänomens sehr gut zeigen können, andererseits aber auch Auskunft über die genaue räumliche Verteilung geben. Ihre Informationsdichte ist deshalb sehr hoch und ihre Produktion, auch mit Computerunterstützung, relativ aufwendig.

Punktkarten arbeiten ebenfalls mit punktförmigen Signaturen, deren Größe jedoch nicht variiert wird. Einem Punkt ist eine feste Größe zugeordnet, z.B. könnte in einer Punktkarte mit dem Thema „Weizenanbau“ ein Punkt einem Hektar Weizenfelder entsprechen. Abb. 2-10 ist ein gutes Beispiel für Punktkarten und zeigt noch einmal ihre beiden Vorteile: Sie zeigen gleichzeitig die Dichte und die Verteilung eines Phänomens.



**Abb. 2-10:** Punktekte [Tyne92]

### Isolinienkarten

Isolinienkarten (engl. isarithmic maps) sind eine Darstellungsform für quantitative Phänomene, die nicht an Punkten oder Flächen aggregiert sind, sondern geographisch kontinuierlich auftreten. Sie werden meistens durch Datensätze repräsentiert, die die Größe des Phänomens an zahlreichen Punkten des Raums wiedergeben. Die Temperatur ist ein solches kontinuierliches Phänomen, und sie kann, z.B. an Meßstationen, nur punktuell erfaßt werden.

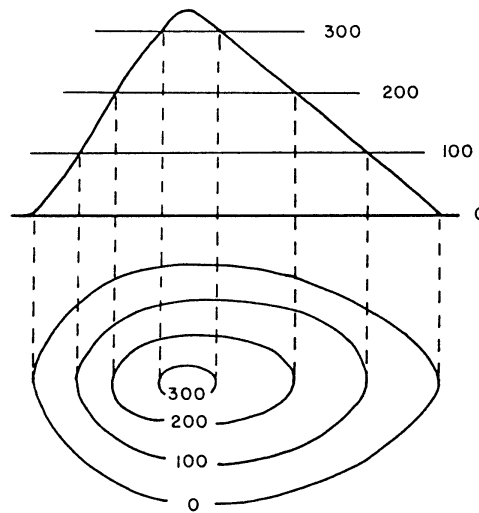
Quantitative, kontinuierliche Phänomene definieren eine sogenannte *statistische Fläche*. Dies ist die Fläche, die entsteht, wenn die Größe des Phänomens an jedem Punkt  $(x,y)$  der geographischen Fläche als Höhe  $z$  interpretiert wird. Die so entstandenen Punkte  $\{(x,y,z)\}$  definieren eine Fläche im dreidimensionalen Raum. Isokarten stellen die statistische Fläche planar dar. Dazu werden in regelmäßigen Abständen horizontale Schnitte durch die Fläche gemacht und die entstandenen Schnittlinien auf die Karte übertragen. Abb. 2-11 zeigt eine statistische Fläche im Querschnitt und das Entstehen einer Isolinienkarte.

Isolinienkarten treten in zwei Ausführungen auf: als reine Isolinienkarten und als Isolinienkarten mit Flächenfüllung. Reine Isolinienkarten zeigen nur die Schnittlinien der statistischen Fläche, die noch durch die Höhe der jeweiligen Schnittebene ergänzt werden müssen (siehe Abb 2-11). Bei Isolinienkarten mit Flächenfüllung wird die Fläche zwischen je zwei Isolinien mit einer Flächensignatur, ganz ähnlich wie bei den Choroplethen ausgefüllt. Eine Legende gibt die Zuordnung von Flächensignaturen zu Höhenintervallen wieder.

### Richtungs- und Flußkarten

Richtungs- und Flußkarten (engl. flow maps) geben Beziehungen zwischen Punkten auf der geographischen Fläche wieder. Sind die Beziehungen gerichtet, werden sie gewöhnlich durch Zeichnen eines Pfeiles zwischen den Punkten symbolisiert. Ist ihnen zusätzlich eine quantitative Größe zugeordnet, z.B. ein Maß für die Stärke des Flusses, wird diese





**Abb. 2-11:** Das Konzept der Isokarte [Tyne92]

Größe auf die Liniendicke des Pfeiles abgebildet. Typische Beispiele für Flußkarten sind Darstellungen von Verkehrsdichten und Handelsbeziehungen.

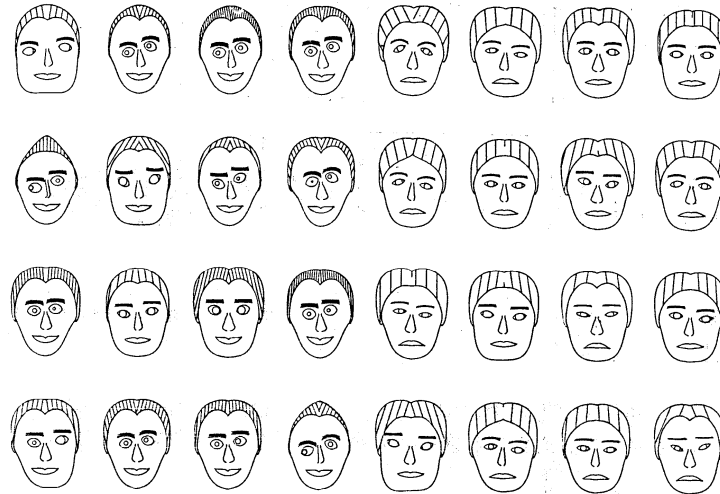
#### **Qualitative Karten (Positions-, Netz- und Mosaikkarten)**

Positions-, Netz- und Mosaikkarten bilden nicht-quantitative Phänomene und Datensätze ab. Sie erreichen dies, indem jedem Wert, der im Datensatz vorkommt, eine geeignete Signatur zugeordnet wird. Je nach der Natur des abgebildeten Phänomens können dies punktförmige, linienförmige oder flächenförmige Signaturen sein. Diese Verfahren sind sehr einfach anzuwenden, da keine Datenklassifikation durchgeführt wird und die Signaturen nicht geordnet oder gestuft sein müssen. Es sollte jedoch beachtet werden, daß die Signaturen das dargestellte Phänomen sinnvoll repräsentieren, indem sie Konventionen der Kartographie und Assoziationen des Kartenlesers berücksichtigt. Beispielsweise sollten bei der Abbildung von Verkehrsnetzen Autobahnen mit der üblichen zweispurigen Liniensignatur und Eisenbahnlinien mit der gestrichelten Liniensignatur dargestellt werden.

Drei Spezialformen der qualitativen Karten treten häufig auf:

- Positionskarten dienen vor allem der Orientierung über die relative Lage und das Vorkommen von Punktphänomenen.
- Netzkarten zeigen die Topologie und die Geographie von Liniennetzen.
- Mosaikkarten bilden eine Unterteilung des geographischen Bereichs in Bezirke, denen jeweils eine qualitative Größe zugeordnet ist, ab. Sie sind das Pendant der Choroplethen für qualitative Daten.

### 2.4.5 Darstellungen von Multiparameterdaten



**Abb. 2-12:** Darstellung multivariater Datensätze mit Gesichtern. Im Bild werden verschiedene Längenmaße echter Banknoten (links) mit denen gefälschter Banknoten (rechts) verglichen [FIRi81].

Einige Darstellungsformen wurden auch aus der Statistik für die Darstellung geographischer Multiparameterdaten übernommen [AnBa96, ScLG96, SyMC96]. Multiparameterdatensätze enthalten eine große Zahl abhängiger Variablen, mindestens 4 oder 5, und können deswegen nicht mit den gewöhnlichen kartographischen Überlagerungen übersichtlich dargestellt werden. Die Multiparametertechniken können dann auf raumbezogene Datensätze angewendet werden, wenn jeweils für einen Punkt auf der geographischen Fläche alle abhängigen Variablen definiert sind und diese Punkte nicht zu dicht zusammenliegen.

Eine bekannte Form der Darstellung von Multiparameterdaten sind die Chernoff-Gesichter. Sie nehmen typischerweise bis zu sechs abhängige Variablen auf, die jeweils auf die Formen und Längen des Gesichts, der Augen, der Nase, des Mundes etc. abgebildet werden. Chernoff argumentierte, daß Menschen in der Betrachtung und im Vergleich von Gesichtern besonders geübt sind und Unterschiede und gemeinsames Auftreten von charakteristischen Attributen gut erkennen und vergleichen können [Cher73, FIRi81]. Chernoff-Faces eignen sich besonders zum Erkennen von Clustern.

Eine andere, etwas einfachere Darstellungsform ist der sogenannte Sunflower-Glyph. Hier werden die Variablen (maximal 6-8) auf die Länge von Linien abgebildet, die um einen Punkt herum angeordnet sind. Es gibt noch eine Anzahl weiterer Techniken, die teilweise auch Farbe, Schriftgröße und Typographie geschickt einsetzen. Allen diesen Techniken ist gemeinsam, daß es gewisse Bedingungen bei der Zuordnung von Datenvariablen zu den Parametern der Techniken gibt. Beim Sunflower-Glyph sollten beispielsweise zwei Variablen, die direkt miteinander verglichen werden müssen, auch benachbarten Linien zugeordnet sein.

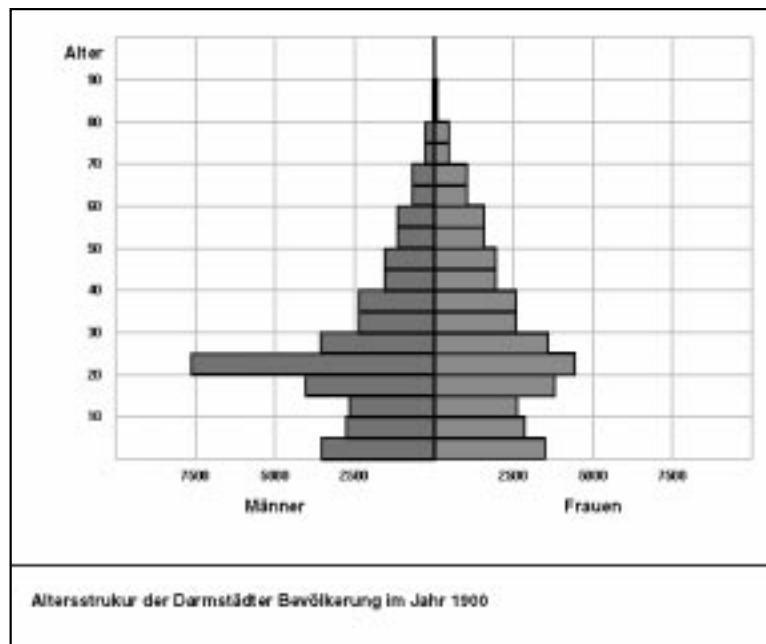
### 2.4.6 Diagramme und Diagrammkarten

Diagramme und statistische Graphen sind Visualisierungsformen für relationale Datensätze ohne Raumbezug. Solche Datensätze treten im GIS zum Beispiel nach Analyse- und Verarbeitungsschritten auf, die einen Datensatz über den geographischen Raum summieren oder mitteln. Für die Diagrammformen gelten jeweils sehr spezifische Regeln, auf welche Art von Datensätzen sie angewendet werden dürfen:

- *Balkendiagramm:* Eine Darstellungsform für Funktionen  $f : A \rightarrow B$ , wobei die Wertemenge  $B$  quantitativ sein muß. Für nominale Wertemengen, z.B. Landnutzungsarten, sind Balkendiagramme nicht geeignet.
- *Scatter-Plot:* Diese Darstellungsform ist allgemeiner verwendbar, sie eignet sich für binäre Relationen  $R : A \times B$  mit beliebigen skalaren Mengen  $A, B$ .
- *Liniengraph:* Diese Darstellungsform ist für Funktionen  $f : A \rightarrow B$  geeignet. Die Menge  $A$  muß geordnet sein,  $B$  quantitativ und zusätzlich stetig.
- *Tortendiagramm:* Eine weitere Darstellungsform für Funktionen  $f : A \rightarrow B$  in quantitative Mengen  $B$ . Zusätzlich muß eine Grundmenge  $G$  gegeben sein, auf der  $A$  eine vollständige Partition definiert.  $f(a)$  ist dann immer die Größe einer Partition von  $G$  und  $\sum_A f(a)$  ist die Größe von  $G$ . Zum Beispiel könnte  $G$  die Menge der gültigen Stimmen in einer Wahl,  $A$  die Menge der zugelassenen Parteien und  $f(a)$  die Anzahl der Stimmen für die Partei  $a$  sein.
- *Boxplot:* Dies ist eine komplexere Form des Balkendiagramms, die sich besonders für die Darstellung statistischer Kenngrößen eignet. Sie zeigt für jede statistische Variable neben dem Median und dem Wertebereich auch die Quartile.
- *Sonderdiagrammformen:* Abhängig vom Anwendungsgebiet des GIS treten auch Sonderdiagrammformen aus den Anwendungsgebieten auf, wie z.B. eine Alterspyramide (Abb. 2-13) in einem GIS für historische Sozialforschung. Für diese Sonderformen gelten ebenfalls spezifische Regeln, auf welche Arten von Datensätzen sie angewendet werden können.

*Diagrammkarten* entstehen, wenn eine der Diagrammformen wiederholt auf einer Karte platziert wird. Zum Beispiel können Tortendiagramme der Wahlergebnisse für verschiedene Wahlbezirke in einer Karte zusammengestellt werden. Das Ergebnis ist eine Tortendiagrammkarte. Grundsätzlich läßt sich jede Diagrammform zur Diagrammkarte verallgemeinern und für Datensätze anwenden, die zusätzlich noch eine unabhängige Variable mit Raumbezug enthalten. Tortendiagrammkarten können also Funktionen der Form  $f : R, A \rightarrow B$  darstellen, wobei  $R$  raumbezogen ist.

Nicht jede der Diagrammformen eignet sich gleich gut für die Verwendung in Karten. Wichtig ist, daß die entstehende Diagrammkarte übersichtlich ist und den Raumbezug bzw. die räumliche Verteilung gut erkennen läßt. Besonders geeignet sind die Tortendia-



**Abb. 2-13:** Sonderdiagrammformen in GIS: eine Alterspyramide

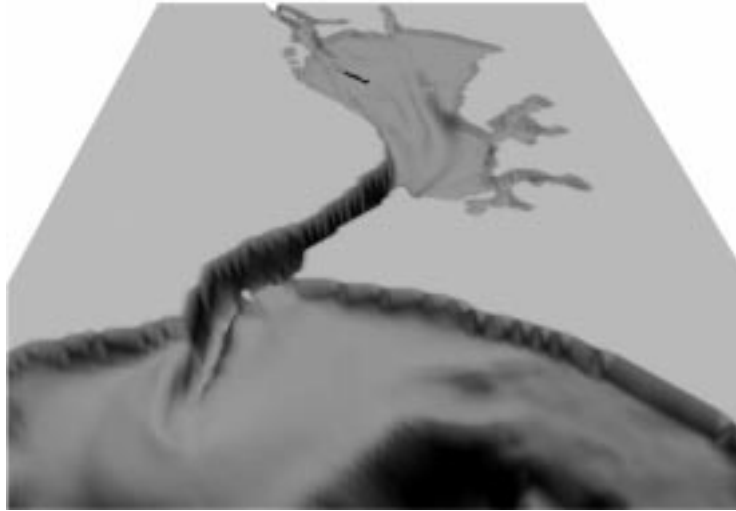
grammkarten, da ihre Kreissignaturen kompakt und einfach strukturiert sind. Liniengraphen und Scatter-Plots wird man dagegen kaum in Karten verwenden.

## 2.4.7 Dreidimensionale Darstellungen

### Perspektivische Terrairndarstellung

Die perspektivische Darstellung von Terrain ist eine naheliegende Anwendung der 3D-Computergraphik in Geo-Informationssystemen. Sie ermöglicht eine neue, dynamische Sichtweise von Geländeformen. Mit modernen Graphikworkstations ist eine Echtzeitdarstellung von komplexem Terrain unter Veränderung der Viewing-Parameter, z.B. ein „virtueller Flug“, realisierbar. Abb. 2-14 zeigt ein Beispiel einer solchen perspektivischen Darstellung, es wird die Meerestiefe der Küste vor Lissabon gezeigt. Höhendaten können heute mit stereoskopischen Luftaufnahmen in hoher Präzision und Auflösung gewonnen werden. Sogenannte Höhenmodelle wandeln diese Höhendaten in ein Polygonnetz für die graphische Repräsentation des Terrains um.

Insgesamt hat die perspektivische Terrairndarstellung im Vergleich zur klassischen, planimetrischen Darstellung mit farbigen Höhenstufen sowohl Vorteile als auch Nachteile, so daß beide Formen nebeneinander bestehen bleiben. Die Terrairndarstellung vermittelt zwar in einigen Fällen einen realistischen, naturgetreuen Eindruck. Häufig müssen die dargestellten Werte jedoch überhöht werden, insbesondere wenn sie nur gering schwanken.



**Abb. 2-14:** Perspektivische Terraindarstellung der Meerestiefe vor Lissabon [Jung93]

### Rendering statistischer Flächen

Ähnlich wie die Höhendaten aus der Natur lassen sich auch statistische Flächen perspektivisch darstellen. Dieses direkte Rendering statistischer Flächen ist gelegentlich als Alternative zur Isoliniendarstellung anzutreffen. Die Vorteile dieser 3D-Visualisierungsform sind ähnlich wie die der Terraindarstellung: sie enthält weniger Abstraktionsschritte als die Isolinkarten, da der Datensatz nicht klassifiziert wird. Damit sind auch kleinere Variationen deutlich sichtbar, die bei Isolinkarten verloren gehen.

Werden statistische Flächen in GIS dreidimensional dargestellt, muß darauf geachtet werden, daß keine Verwechslungen mit Höhendaten möglich sind. Insbesondere Betrachter, die solche Visualisierungstechniken für statistische Flächen nicht kennen, neigen dazu, sie als Geländeform zu interpretieren. Diese Tendenz wird noch verstärkt, wenn naturnahe Farbtöne, z.B. Grün und Braun, in der Visualisierung verwendet werden.

### Virtuelle Städte und Landschaften

Ein wichtiges Anwendungsgebiet der Virtuellen Realität liegt in der realistischen und dynamischen Visualisierung von Städten und Landschaften. Dazu wird ein Höhenmodell der Landschaft mit einem Architekturmodell einer Stadt kombiniert. Virtuelle Stadt- und Landschaftsmodelle dienen vor allem der Evaluierung von Planungen, z.B. von Neubauten oder Veränderungen im Stadtbild.

## 2.4.8 Visualisierung von Vektorfeldern

Die Mehrzahl der Techniken der technisch-wissenschaftlichen Visualisierung sind direkt auf Geodaten anwendbar und werden deshalb immer häufiger auch im GIS-Umfeld eingesetzt. Es kann hier nicht auf jede einzelne dieser Methoden eingegangen werden, statt

dessen sei auf die wissenschaftliche Literatur zu diesem Thema verwiesen [Brod92, KeKe92].

Eine Klasse dieser Visualisierungstechniken wird insbesondere auch im GIS-Kontext verstärkt eingesetzt: die Techniken zur Visualisierung von Vektorfeldern. In Umweltinformationssystemen spielt die Modellierung und Visualisierung von Strömungen und Flüssen eine wichtige Rolle. Sie werden in Form zwei- oder dreidimensionaler Vektorfelder modelliert und mit verschiedenen Techniken visualisiert. Man unterscheidet zwischen Techniken, die die Strömungsfelder direkt visualisieren und solchen, die sie indirekt, zum Beispiel über ihre Wirkung auf Partikel oder Gegenstände darstellen. Die vier wichtigsten Techniken sind:

- *Pfeildarstellungen:* Diese Technik zeigt das Vektorfeld direkt, jeder Vektor wird durch einen gerichteten Pfeil repräsentiert, die Vektorintensität kann z.B. farbkodiert sein. Pfeildarstellungen sind insbesondere für 2D-Felder oder für Schnitte in 3D-Feldern geeignet (siehe Abb 2-15)
- *Partikelverfolgung:* Es werden (masselose) Teilchen im Raum plaziert und die Auswirkung des Vektorfeldes auf die Teilchen simuliert. Das Ergebnis ist eine Animationssequenz, in der die Bewegung der Teilchen zu sehen ist.
- *Partikelbahnen:* Es werden ebenfalls Teilchen verfolgt, die Teilchenbahnen werden festgehalten und als Röhren gerendert. Die Geschwindigkeit kann farbkodiert werden. Das Ergebnis ist ein Bild der Partikelbahnen (siehe Abb 2-5).
- *Papierstreifen* (flow ribbons): Die Bewegung von fest verankerten Streifen („Papierstreifen“) wird simuliert und gerendert.

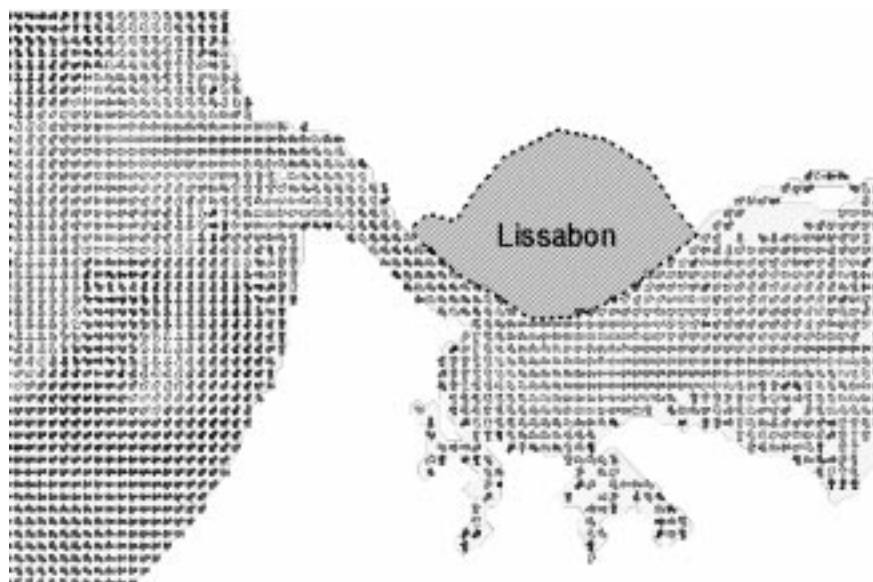


Abb. 2-15: Pfeildarstellung eines Strömungsfeldes [Jung93]

### 2.4.9 Animation

Gleichzeitig mit der dritten Dimension hält auch die vierte Dimension, die Zeit, schrittweise Einzug in die Geo-Informationssysteme. Die Schwierigkeiten, die die Aufnahme der Zeit für die Datenmodellierung, die Datenhaltung und die Interaktion in den GIS bereitet, sind bis heute nur teilweise gelöst und temporale GIS bisher nur als Prototypen anzutreffen [Lang92, JGLT95]. Interessante Resultate gibt es bei der Visualisierung zeitabhängiger, raumbezogener Daten. Hier wurden in den vergangenen Jahren einige theoretische und praktische Fortschritte gemacht. Inspiriert von der Dynamik des neuen Kartenmediums Computer wurden die vielfältigen Einsatzmöglichkeiten der kartographischen Animation, vor allem in der Ausbildung, demonstriert [Monm92]. MacEachren und Kraak haben die verschiedenen Möglichkeiten, Zeit als kartographische Variable einzusetzen, untersucht und bewertet [MacE94b, KrMa94].

#### Zeit als kartographische Variable

Vergleicht man statische Karten mit animierten Karten bezüglich der Ausdrucksmittel oder der visuellen Variablen, so könnte man erwarten, daß in animierten Karten zu den Variablen Position, Farbe, Größe, etc. eine einzige neue Variable hinzutritt: die (Darstellungs-)Zeit. Untersuchungen zeigen jedoch, daß die Zeit nicht als eine einzige, homogene visuelle Variable wahrgenommen wird, sondern daß sie bis zu sechs neue Variablen induziert. Kraak und MacEachren nennen diese sechs Variablen Darstellungszeit, Dauer, Frequenz, Reihenfolge, Änderungsrate und Synchronisation [KrMa94]. Mit der Variable „Dauer“ kann beispielsweise ausgedrückt werden, wie lange eine Signatur auf einer Karte zu sehen ist, mit „Frequenz“ die Blinkrate einer blinkenden Signatur ausgedrückt werden. Diese sechs neuen Variablen haben naturgemäß sehr verschiedene charakteristische Eigenschaften und sie können auch nicht alle gleichzeitig eingesetzt werden, ohne die Überschaubarkeit der Darstellung zu opfern.

#### Animation dreidimensionaler Darstellungen

Diese visuellen Variablen können nicht nur für planimetrische kartographische Darstellungen sondern auch für dreidimensionale Visualisierungen eingesetzt werden. Die visuelle Variable „Darstellungszeit“ ist im 3D-Bereich jedoch dominierend, die anderen fünf Variablen werden nur vereinzelt gezielt eingesetzt. Drei Arten der Verknüpfung der Darstellungszeit mit Parametern der Visualisierung sind möglich:

- *Modellzeit = Darstellungszeit:* Es wird ein zeitabhängiger Datensatz oder eine zeitabhängige Simulation visualisiert. Der Zeitparameter des dargestellten Datensatzes ändert sich dabei proportional zur Darstellungszeit.
- *Modellparameter = Darstellungszeit:* Eine andere Dimension (bzw. unabhängige Variable) des Datensatzes wird proportional zur Darstellungszeit verändert.
- *Kamerafahrten:* Die Viewing-Parameter (z.B. Kameraposition, -orientierung) einer perspektivischen Darstellung werden abhängig von der Darstellungszeit verändert.

## 2.5 Effektive GIS-Visualisierung

Im letzten Abschnitt wurde gezeigt, wie vielfältig die Techniken der Visualisierung in modernen GIS sein können. An verschiedenen Stellen wurde auch schon deutlich, daß für die Anwendbarkeit vieler Techniken bestimmte Bedingungen an den Datensatz oder an das Umfeld der Visualisierung erfüllt sein müssen. Abschnitt 2.5 baut diese Beobachtung aus und stellt sie auf eine theoretische Grundlage: die Semiotik der graphischen Darstellungen. Die zentralen Begriffe der Ausdrucksfähigkeit und der Effektivität graphischer Darstellungen werden definiert und beispielhaft einige Regeln aus der Kartographie und der Computervisualisierung vorgestellt.

### 2.5.1 Ausdrucksfähigkeit und Effektivität von Visualisierungen

#### Ausdrucksfähigkeit

Die zentrale Forderung an Visualisierungen von Datensätzen ist die Ausdrucksfähigkeit. Visualisierungen sollen einen Datensatz möglichst unverfälscht wiedergeben, sie sollen ihn *ausdrücken* können. Ausdrucksfähigkeit kann, angelehnt an [Mack86b] folgendermaßen definiert werden:

##### Definition 2-1 (Ausdrucksfähigkeit)

Eine Visualisierung heißt *ausdrucksfähig* für einen Datensatz, wenn sie

- (1) *jede* Information des Datensatzes enthält und
- (2) *ausschließlich* die Information des Datensatzes enthält.

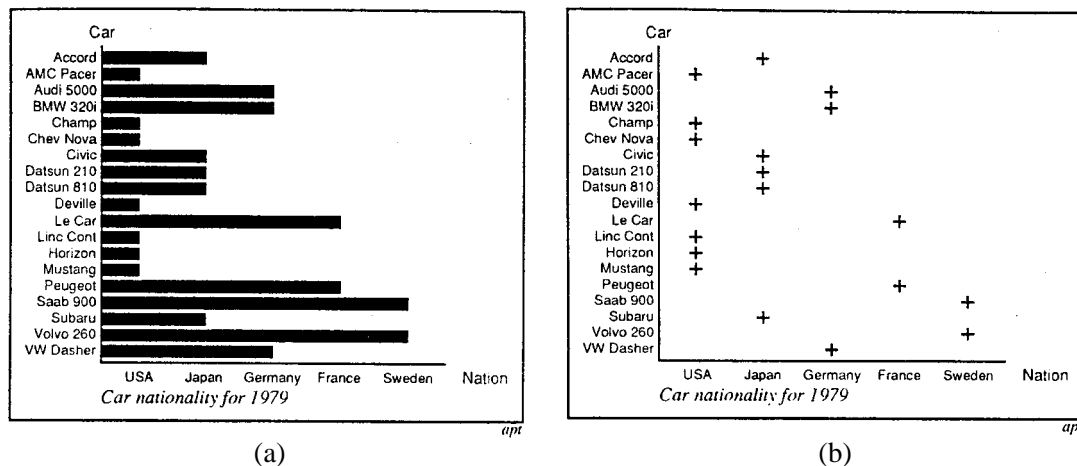
Im Gegensatz zu anderen, insbesondere künstlerischen Arten graphischer Darstellung, tritt bei der Visualisierung die Funktion gänzlich vor die Form. Alle Arten von Ornamenten und Ausschmückungen sind deswegen im Laufe der Zeit aus Visualisierungen und Karten fast vollständig verdrängt worden. Tuft demonstriert eindringlich, wie die Reduktion auf die wesentliche Elemente den „Gebrauchswert“ graphischer Darstellungen steigert [Tuft83, Tuft90].

Die Ausdrucksfähigkeit kann aber auch auf subtilere Weise verletzt werden, wie die beiden Diagramme in Abb. 2-16 zeigen. Beide Diagramme versuchen, dieselbe Information darzustellen: eine Zuordnung von Automarken zu Herstellerländern. Aber nur der Scatter-Plot (b) ist ausdrucksfähig. Im Balkendiagramm (a) suggeriert die unterschiedliche Länge der Balken eine Ordnung der Länder derart, daß beispielsweise Schweden größer oder besser sei als die USA. Da eine solche Information über die Herstellerländer in dem Ausgangs-Datensatz jedoch nicht enthalten ist, muß (a) als nicht ausdrucksfähig eingestuft werden.

#### Effektivität

Die zweite zentrale Forderung an Visualisierungen ist die Effektivität. Wenn mehrere Visualisierungstechniken einen Datensatz ausdrücken, will man natürlich diejenige finden, die ihn am besten ausdrückt. Im Gegensatz zur Ausdrucksfähigkeit hängt die Effektivität einer Visualisierung außer vom Datensatz noch von anderen Einflußfaktoren ab. Insbe-





**Abb. 2-16:** Zwei Visualisierungen derselben Relation. Der Scatter-Plot (b) ist ausdrucksfähig, das Balkendiagramm (a) ist nicht ausdrucksfähig. [Mack86b]

sondere hängt sie ab von der menschlichen visuellen Wahrnehmung und den spezifischen visuellen Fähigkeiten des Betrachters der Visualisierung. Die Effektivität einer Visualisierung kann aber auch mit den Eigenschaften des Ausgabemediums variieren, zum Beispiel könnte eine Visualisierung auf dem Bildschirm effektiver sein als auf Papier.

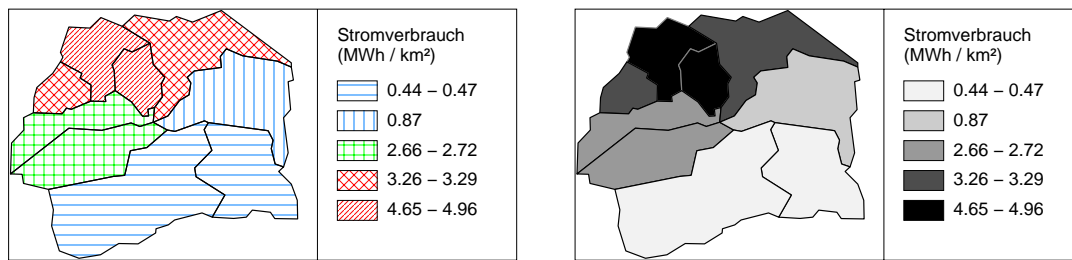
Die folgende Definition der Effektivität berücksichtigt ihre Einflußfaktoren:

#### Definition 2-2 (Effektivität)

Eine Visualisierung ist *optimal effektiv*, wenn sie für die Darstellung eines Datensatzes die visuell-kognitiven Fähigkeiten des Betrachters und die Möglichkeiten des Ausgabemediums für das Erreichen der Visualisierungszielsetzung optimal ausnutzt.

Ein weiterer Einflußfaktor ist in dieser Definition enthalten: die Visualisierungszielsetzung. Visualisierungen können verschiedenen Aufgaben dienen, sie können beispielsweise dazu dienen, bestimmte Werte in einem Datensatz möglichst leicht zu finden, oder aber den Vergleich aller Werte des Datensatzes zu vereinfachen. Eine Visualisierung, die für die erste Aufgabe effektiv ist, kann für die zweite Aufgabe versagen, und umgekehrt.

Abb. 2-16 zeigt ein Beispiel zur Illustration von Effektivität. In beiden Fällen wird eine quantitative, raumbezogene Variable durch eine Choroplethenkarte dargestellt. Die verwendete visuelle Variable ist unterschiedlich: einmal ist es „Textur“, das andere Mal „Helligkeitswert“. Offensichtlich ist in der rechten Karte der Wertevergleich einfacher, da die verschiedenen Helligkeitswerte vom Betrachter spontan in die richtige Reihenfolge gebracht werden können. Bei den unterschiedlichen Schraffuren in der linken Karte ist das nicht so ohne weiteres möglich. Vergleicht man in dieser Karte Werte, so muß gelegentlich in der Legende nachgesehen werden, die Karte ist also für den Wertevergleich weniger effektiv.



**Abb. 2-17:** Zwei Darstellungen einer quantitativen, raumbezogenen Variablen. Die linke ist weniger effektiv als die rechte.

### 2.5.2 Elemente der Graphischen Semiotik

Warum ist eine der Karten in Abb. 2-17 offensichtlich effektiver als die andere? Der französische Kartograph Jacques Bertin hat sich in den sechziger Jahren intensiv mit dieser Frage beschäftigt und als Antwort eine Theorie der Graphischen Darstellungen aufgestellt, in der sich dieses und viele ähnliche Phänomene erklären lassen [Bert74]. Seine Theorie, die Graphische Semiotik (bzw. Graphische Semiologie) enthält unter anderem eine Methodik für die Bestimmung ausdrucksfähiger und effektiver Visualisierungen eines Datensatzes.

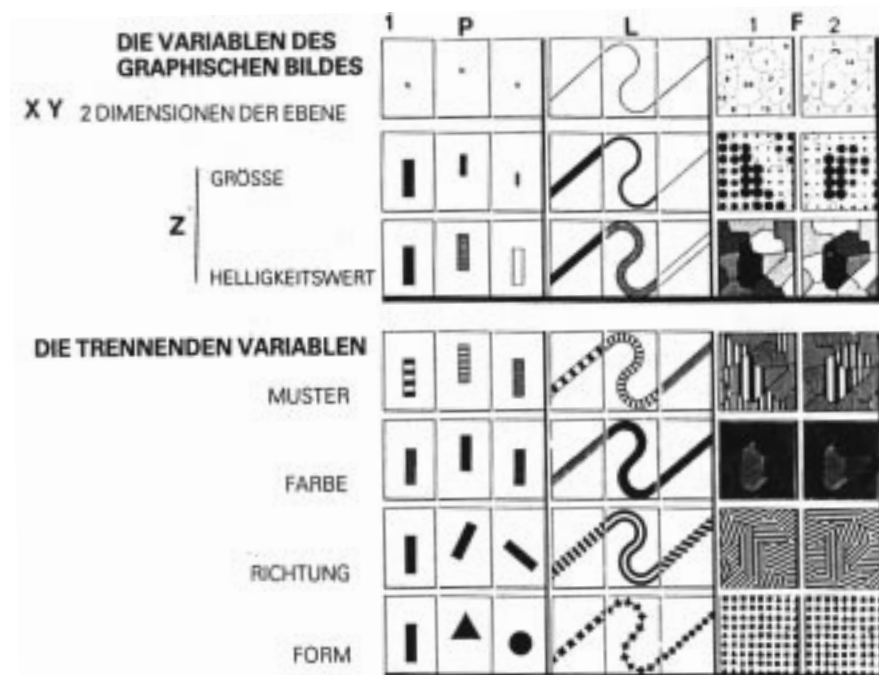
#### Datencharakteristika

Bertins Methodik beginnt mit einer Charakterisierung des Datensatzes, der dargestellt werden soll. Er unterscheidet die verschiedenen Komponenten (auch Variablen oder Attribute genannt) eines Datensatzes und betrachtet jeweils ihre Charakteristika. Das wichtigste Charakterisierungsmerkmal der Komponenten ist ihr Ordnungstyp: Eine Komponente kann nominal, ordinal, quantitativ oder geographisch sein. Außerdem müssen die Länge und die Anzahl der Komponenten betrachtet werden. Die Stromverbrauchsrelation enthält beispielsweise zwei Komponenten: eine geographische (der Ort) und eine quantitative (der Verbrauch). Nominale Komponenten sind ungeordnet (z.B. Automarken). Ordinale Komponenten sind geordnet, aber ihnen entsprechen keine präzisen Zahlenwerte (z.B. sind die verschiedenen Modelle eines Autoherstellers von „Kleinwagen“ bis „Limousine“ geordnet).

#### Visuelle Variablen

Bertins Lehre der graphischen Semiotik katalogisiert die Ausdrucksmittel graphischer Darstellungen. Er identifiziert acht Dimensionen oder visuelle Variablen, die unabhängig voneinander variiert werden können: Größe, Helligkeit, Muster, Farbe, Richtung und Form, sowie die zwei räumlichen Dimensionen der Ebene (siehe Abb 2-18). Jede dieser Variablen unterstützt bestimmte Formen graphischer Wirkung. Bezogen auf die Darstellung von Information werden vor allem drei Wirkungsformen unterschieden:

- *selektiv:* Signaturen werden vom Betrachter spontan in Gruppen aufgeteilt. Die Variable eignet sich zur Darstellung nominaler Daten. Beispiel: Landnutzungsarten werden durch Farbtöne symbolisiert, wie Wald durch Grün, geschlossene Siedlungen durch Rot und Ackerland durch Gelb.



**Abb. 2-18:** Die acht visuellen Variablen [Bert82]

- *ordinal:* Signaturen werden vom Betrachter spontan in eine Reihenfolge gebracht. Die Variable eignet sich zur Darstellung ordinaler Daten. Beispiel: Die Waldschadensklassen I-IV werden durch unterschiedliche Intensitäten der Farbe Grün dargestellt.
- *proportional:* Signaturen werden vom Betrachter ebenfalls in eine Reihenfolge gebracht. Zusätzlich kann leicht ein Datenwert proportional zum Wert der visuellen Variable assoziiert werden. Die Variable eignet sich zur Darstellung ordinaler und quantitativer Daten. Beispiel: Einwohnerzahlen von Städten werden durch proportional große Kreissignaturen dargestellt.

Tabelle 2-1 zeigt, welche Wirkung die verschiedenen visuellen Variablen haben können und zur Darstellung welcher Datenvariablen sie folglich geeignet sind. Durch diesen Zusammenhang zwischen Ordnungstyp der Datenvariablen und Eignung visueller Variablen wird auch klar, warum die Effektivität der Karten in Abb. 2-17 unterschiedlich ist. Die quantitative Komponente „Stromverbrauch“ wird einmal mit der geeigneten visuellen Variablen „Helligkeitswert“ symbolisiert und es resultiert eine effektive Darstellung. Im anderen Fall wird sie mit der ungeeigneten visuellen Variablen „Muster“ symbolisiert und es resultiert eine ineffektive Darstellung.

Visuelle Variable	Wirkung		
	selektiv	ordinal	proportional
Dimensionen der Ebene	✗	✗	✗
Größe	✗	✗	✗
Helligkeitswert	✗	✗	✗
Muster	✗	✗	
Farbe	✗		
Richtung	✗ (nicht bei Flächen)		
Form			

Tabelle 2-1: Wirkung der visuellen Variablen

### 2.5.3 Der Entwurf thematischer Karten

Ausdrucksvolle und effektive Darstellungen sind das Ziel des Visualisierungsentwurfs. Die graphische Semiotik schreibt die Grundsätze des Entwurfs von Darstellungen vor. In vielen Bereichen der graphischen Kommunikation von Information gibt es jedoch darüber hinaus noch detaillierte Regelwerke, die die Wahl und Ausgestaltung der Visualisierungstechniken bestimmen. Im Falle der (thematischen) Kartographie sind diese Regelwerke besonders umfangreich. Dies liegt zum Teil an der langen Tradition der Kartographie, deren Techniken im Laufe der Jahrhunderte immer weiterentwickelt wurden und heute sehr ausgereift sind. Ein weiterer Grund ist aber sicherlich die Komplexität der Darstellung raumbezogener Information selbst: sie bringt viele Schwierigkeiten mit sich, die bei anderen, nicht-raumbezogenen Datensätzen nicht vorkommen, z.B. muß die Erdoberfläche zuerst projiziert werden, bevor sie in einer planimetrischen Karte abgebildet werden kann.

#### Vorverarbeitung

Der Entwurf einer thematischen Karte beginnt deswegen mit der Auswahl einer geeigneten Projektionsmethode für den abzubildenden geographischen Raum. Um eine geeignete Grundkarte für die thematische Karte zu erhalten, müssen prinzipiell alle weiteren kartographischen Entwurfsschritte, insbesondere die Generalisierung, durchlaufen werden. Häufig kann an dieser Stelle jedoch auf eine bereits existierende Grundkarte, z.B. eine topographische Karte zurückgegriffen werden, die für die Aufnahme zusätzlicher thematischer Information in hellen Grautönen gehalten ist.

Im nächsten Schritt erfolgt eine Vorverarbeitung des darzustellenden Datensatzes. Die relevante Information muß ausgewählt werden und bei einem Maßstabswechsel kann auch eine Generalisierung des raumbezogenen Datensatzes notwendig sein.

### **Symbolisierung**

Im Symbolisierungsschritt werden für den Datensatz geeignete, d.h. ausdrucksfähige und effektive Kartentypen bestimmt. Es gibt für alle thematisch-kartographischen Techniken umfangreiche Regelwerke, die ihre Ausdrucksfähigkeit und Effektivität beschreiben, z.B. [Dent96, Tyne92]. Die Regeln basieren immer auf den Gesetzen der graphischen Semiotik, ergänzen diese aber um zahlreiche weitere Kriterien und Bedingungen. Manche dieser Regeln sind offensichtlich, andere werden jedoch häufig übersehen und führen beim naiven Anwender immer wieder zu fehlerhaften Darstellungen. Im nächsten Abschnitt werden Beispiele solcher Darstellungen gezeigt.

### **Parameterwahl**

Parameter thematischer Karten können einfache und gestufte Signaturen, sowie Klassen- aufteilungen sein. Eine Datenklassifikation ist bei den meisten quantitativen Methoden erforderlich. Die Klassifikation muß unbedingt methodisch durchgeführt werden und den Datensatzcharakteristika entsprechen, damit die resultierende Karte aussagekräftig ist. Auch für die Auswahl der Signaturparameter einer thematischen Karte gibt es feste Regeln. Bei Signaturkarten sind z.B. in der Regel Kreisscheiben vorzuziehen, da sie sich leichter vergleichen lassen als andere Signaturformen.

### **Visuelle Gestaltung**

Insgesamt muß auf eine ausgewogene und effektive visuelle Gestaltung der Karte geachtet werden. Farbe sollte zielgerichtet eingesetzt sein: Intensive Farben dürfen nur zur Hervorhebung und nur an wenigen Stellen erscheinen; als Flächenfarbe sollten ungesättigte Farben dienen. Ähnliche Regeln gelten auch für Typographie, Legende und den Kontrast der Karte.

## **2.5.4 Methodik der praktischen Computervisualisierung**

Die Computervisualisierung ist eine relativ junge Fachrichtung und ihre Regelwerke sind entsprechend einfacher und weniger ausgereift. Jedoch erschwert hier die Vielzahl der verfügbaren Visualisierungstechniken den Überblick: Wehrend und Lewis kamen bei ihrer Klassifikation [WeLe90] schon im Jahre 1990 auf über 400 Visualisierungstechniken und es kommen jedes Jahr etliche neue dazu. Offensichtlich ist bei einer solchen unüberschaubaren Menge von Techniken eine sinnvolle Auswahl mit der Trial-and-Error-Methode nicht mehr gewährleistet und folgerichtig wurde bald eine Methodik für den Visualisierungsentwurf gefordert. Die Entwicklung dieser Methodik hat in den letzten Jahren einige Fortschritte gemacht, sie ist aber insgesamt noch nicht als ausgereift zu bezeichnen. Parallel wurde auch versucht, den Visualisierungsentwurf zu automatisieren und die Methodik in Form wissensbasierter Systeme zu implementieren (siehe Kapitel 3).

Der systematische Visualisierungsentwurf geht in ähnlichen Schritten vor wie der Entwurf thematischer Karten. Grundlage ist eine Charakterisierung des Datensatzes, die Punkt-, Skalar-, Vektor- und Tensordaten, abhängige und unabhängige Parameter, Dimension und Ordnungstyp der Parameter unterscheidet. In Tabellen und Referenzwerken, z.B. [Brod92], können Visualisierungstechniken, die für einen Datensatz anhand seiner Charakteristika prinzipiell geeignet sind, nachgeschlagen werden. In einem weiteren Schritt muß für jede so gefundene Technik ihre Effektivität bezüglich der Visualisierungszielsetzung und dem Kontext der Visualisierung geprüft werden. Jede der Visualisierungstechniken hat spezifische Vor- und Nachteile, die teilweise in der Originalliteratur nachgeschlagen werden müssen, teilweise aber auch schon in Referenzwerken zusammengestellt sind, so in [Selg90b] und [KeKe92].

Neuerdings wird versucht, die Visualisierungszielsetzung und den Visualisierungskontext in der Methodik stärker zu berücksichtigen. Für die Visualisierungszielsetzung ist es allerdings schwierig, ein allgemein verwendbares und trotzdem genügend detailliertes Charakterisierungsschema zu finden. Erste Ansätze sind die Arbeiten von Wehrend [WeLe90] und Robertson [Robe91], auf die in Abschnitt 3.3 noch genauer eingegangen wird. Von Wehrend stammt auch eine Tabelle, in der die Zielsetzung direkt als Index für die Auswahl einer Visualisierungstechnik dient. Keller und Keller [KeKe92] haben sich bemüht, das Ausgabemedium als wichtigen Aspekt des Visualisierungskontextes beim Visualisierungsentwurf stärker zu berücksichtigen.

## 2.6 Das Problemfeld GIS-Visualisierungsentwurf

In den letzten beiden Abschnitten wurde gezeigt, wie umfassend und komplex die Methoden zur Visualisierung in GIS heute sind, und daß für die Erstellung aussagekräftiger und effektiver GIS-Visualisierungen Kenntnisse der graphischen Semiotik, der Kartographie und der Computervisualisierung erforderlich sind. Visualisieren heißt immer auch Entwerfen: es ist ein kreativer und wissensintensiver Prozeß. Was aber passiert, wenn dieses Visualisierungswissen bei Anwendern von GIS nicht vorhanden ist? Dieser Abschnitt zeigt die resultierenden Probleme auf, gibt Beispiele für Fehlerquellen bei der GIS-Visualisierung und diskutiert mehrere Lösungsansätze.

### 2.6.1 Wissenslücken bei der GIS-Visualisierung

Heutige GIS stellen eine große Zahl von Visualisierungstechniken zur Verfügung. Was in diesen Systemen jedoch fehlt, ist eine geeignete Repräsentation des graphischen Wissens, das zu diesen Techniken gehört und das für ihren sinnvollen Einsatz notwendig ist [WeBu92]. Die GIS-Softwarekomponente für die Visualisierung wurde sehr treffend als funktionsreich und wissensarm charakterisiert [YaST93].

In der Anfangszeit von GIS hatte noch die Mehrzahl der GIS-Anwender eine fundierte kartographische Ausbildung. Heute, mit der Verbreitung von GIS in den verschiedensten Anwendungsbereichen, stammen viele der GIS-Anwender aus Fachgebieten, wo die Kar-

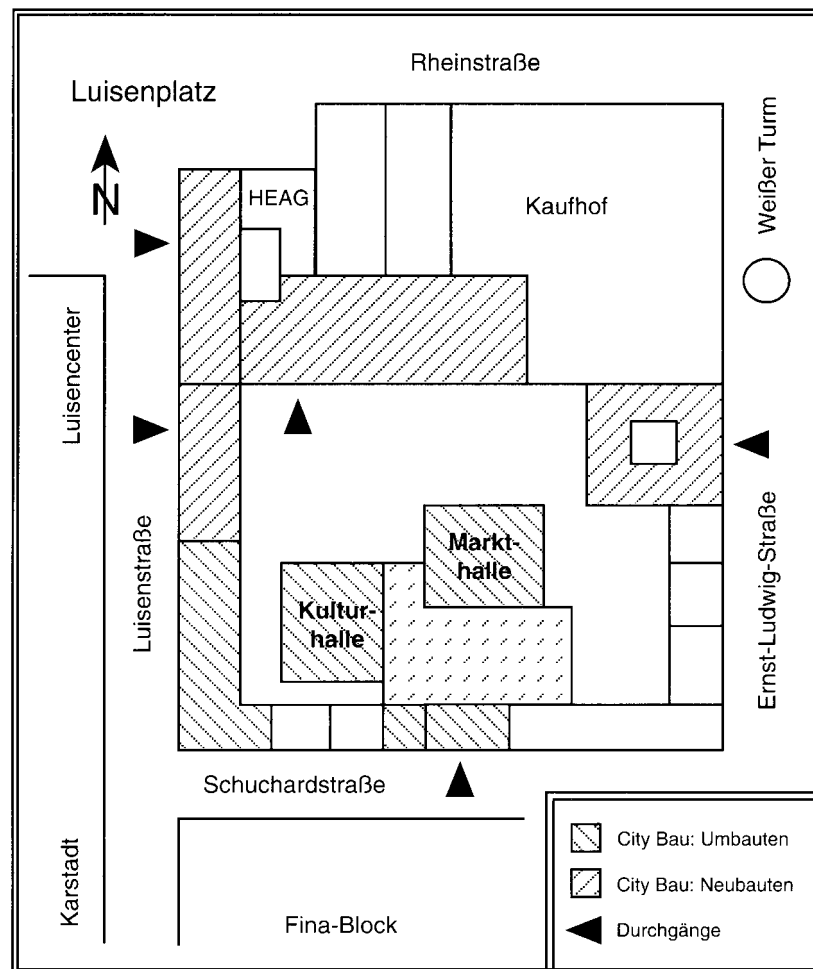
tographie nicht Bestandteil der Ausbildung ist. Sie haben sich ihre kartographischen Kenntnisse durch Beobachtung und Experimente angeeignet und kennen häufig fundamentale Gesetze der graphischen Kommunikation nicht. Trotzdem ermöglichen die GIS es ihnen, eindrucksvolle Visualisierungen und Karten zu erstellen. Diese Visualisierungen sehen zwar professionell aus, erfüllen aber häufig ihren Zweck, die effektive Kommunikation raumbezogener Information, nur unzureichend. Dies kann sogar soweit gehen, daß die ursprüngliche Information in der Darstellung vollkommen verfälscht wiedergegeben wird. In manchen Fällen haben die Graphiken zwar nur wenig Information auszudrücken, tun dies aber mit pompösen Mitteln. Tufté hat eine Reihe solcher peinlicher Graphiken, die er *graphic ducks* nennt, zusammengestellt [Tuft83].

### 2.6.2 Häufige Fehler

In wichtigen Medien wie Tageszeitungen, Magazinen, Büchern und im Fernsehen ist heute eine Flut von Graphiken und Karten zu verzeichnen. Auch bei den raumbezogenen Graphiken ist ein beträchtlicher Anteil fehlerhaft, ein weiterer Beleg dafür, daß die Wissensarmut der Visualisierungstools ins Gewicht fällt. Dieser Abschnitt nennt die häufigsten Fehlerquellen, und zeigt an einigen Beispielen die Auswirkung von Visualisierungsfehlern auf.

Ein Fehler, der bei kartographischen Darstellungen häufig zu sehen ist, ist die Verwendung ineffektiver visueller Variablen. In Abbildung 2-20 sollen Neubauten von renovierten Gebäuden unterschieden werden. Dies wird durch verschieden orientierte Schraffuren der Flächen in der Karte versucht. Eine der Regeln der graphischen Semiotik lautet aber, daß die Orientierung von Signaturen nicht zur Unterscheidung von Flächen verwendet werden kann. Diese Regel läßt sich am Beispiel der Karte deutlich nachvollziehen: die zwei Gebäudetypen sind spontan kaum zu unterscheiden, es muß immer wieder in der Legende nachgesehen werden. Hätte der Kartenautor hier beispielsweise verschiedene Graustufen eingesetzt, wären die Gebäudetypen spontan unterscheidbar. Diese Karte enthält noch einen weiteren Mangel: eines der Gebäude ist gestrichelt schraffiert, aber die Bedeutung dieser besonderen Signatur sucht man in der Legende vergebens.

Vielleicht noch gravierender sind Fehler in der Ausdrucksfähigkeit einer Darstellung. Das Beispiel in Abb. 2-20 gibt ein verfälschtes Bild der zugrundeliegenden Information wieder. In einer Benutzerumfrage konnten verschiedene Online-Dienste mit Noten von 1 (sehr gut) bis 6 (sehr schlecht) bewertet werden. In dem Diagramm werden die Durchschnittsnoten aus der Befragung dargestellt, es ergibt sich aber ein irreführender Eindruck: höhere Noten, also schlechtere Bewertungen werden durch weiter oben liegende Punkte und Linien repräsentiert. In solchen Diagrammen erwartet der Betrachter jedoch, daß die weiter oben liegenden Symbole besseren Bewertungen entsprechen. Der Eindruck, den das Diagramm vermittelt, ist deswegen falsch. Das Diagramm ist aber noch aus einem weiteren Grund nicht ausdrucksfähig: die einzelnen Bewertungen im Diagramm sind durch Linien verbunden. Dadurch entsteht der irreführende Eindruck, die verschiedenen Kriterien wie Handhabung, Angebotsbreite, etc. ergäben eine feste (z.B. zeitliche) Reihenfolge und es sei beispielsweise die zeitliche Entwicklung der Bewertung dargestellt.

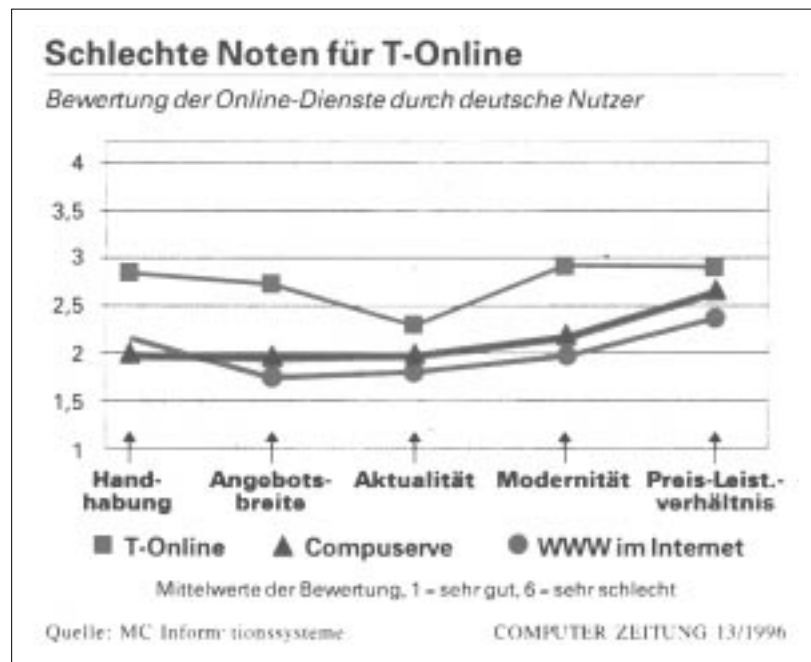


**Abb. 2-19:** Beispiel einer uneffektiven thematischen Karte: durch die Verwendung der unpassenden visuellen Variable *Orientierung* können Umbauten von Neubauten nur schwer unterschieden werden.

Eine nicht ausdrucksfähige Darstellung raumbezogener Information zeigte schon Abb. 1-2: die Choroplethenkarte eines Datensatzes mit Zahlen über den gesamten monatlichen Stromverbrauch in verschiedenen Bezirken. Da es sich dabei um Absolutzahlen und nicht um flächennormierte oder von der Fläche unabhängige Werte handelte, war die Choroplethentechnik nicht angemessen. Die Absolutzahlen hätten beispielsweise mit Kreissignaturen in ausdrucksfähiger Weise dargestellt werden können.

Als problematisch sind auch solche Karten und Visualisierungen einzustufen, die zu aufwendig oder zu komplex sind und deswegen schwierig zu interpretieren sind. Durch die Möglichkeit, sehr aufwendige und professionell erscheinende Graphiken in einem GIS zu produzieren, werden manchmal auch Kartenautoren verführt, die nicht viel Information auszudrücken haben. Auffällige graphische Elemente wie 3D-Effekte, Schattierungen, gesättigte Farben oder eine zu detaillierte Basiskarte tragen häufig nicht dazu bei, den Datensatz auszudrücken, sondern sind reine Gestaltungselemente. Tufte nennt solche Dar-





**Abb. 2-20:** Ein Diagramm mit mehreren Fehlern in der Ausdrucksfähigkeit

stellungen *self-promoting graphics* und zeigt ein Beispiel für die Ineffektivität einer übertrieben gestalteten Karte [Tuft83, S. 119].

Manchmal sind komplexe Darstellungsmethoden zwar durchaus angemessen für einen Datensatz, ein ungeschulter Betrachter ist aber mit der Darstellung überfordert und kann sie nicht interpretieren. Ein Beispiel für solche komplexen Darstellungsmethoden sind die bivariaten Choroplethenkarten, die nicht eine, sondern zwei quantitative Variablen ausdrücken. Ihre zweidimensionale gestufte Flächensignatur wird von vielen Kartographen als für die Kommunikation von Information zu kompliziert eingestuft. Ein Beispiel bringt wieder Tuft [Tuft83, S. 153]. In bestimmten Fällen, z.B. bei der Analyse eines Datensatzes können sie aber durchaus sinnvoll sein. Es ist deswegen immer erforderlich, die Aufgabenstellung und die Zielgruppe einer Visualisierung zu beachten.

### 2.6.3 Lösungsansätze

Wie kann diese Lücke graphischen Wissens in GIS-Anwendungen geschlossen und damit fehlerhafte und ineffektive GIS-Visualisierung vermieden werden? Es gibt zwei prinzipiell unterschiedliche Lösungsansätze für dieses Problem. Zum einen kann man versuchen, die Wissenslücke auf der Anwenderseite zu schließen und möglichst viele potentielle Benutzer von GIS in graphischem Entwurf und in der graphischen Kommunikation von Information ausbilden. Die Alternative dazu ist, die Wissenslücke auf der Systemseite zu schließen und die GIS-Visualisierungskomponente mit graphischem Wissen auszustatten.

Die erste Möglichkeit, also eine breit angelegte Visualisierungs-Ausbildung, hat den Vorteil, daß die Anwender dieses neu erworbene Wissen auch in zahlreichen anderen Bereichen anwenden können und daß sie unabhängig vom Wissensniveau der GIS-Systeme werden. Sie hat allerdings den Nachteil, daß der Lernprozeß naturgemäß Zeit und Kosten erfordert und die Entscheidung für Ausbildungsmaßnahmen erst beim einzelnen Anwender, seinem Arbeitgeber oder übergreifend durchgesetzt werden müßte.

Vorteil der zweiten Möglichkeit, also der Schaffung wissensbasierter Visualisierungssysteme ist, daß das Wissen genau an der Stelle zur Verfügung steht, wo es benötigt wird und damit beim Einzelnen oder der Institution keine Ausbildungsmaßnahmen vorgeschaltet werden müssen. Es hat den Nachteil, daß Computer-Darstellungen menschlichen Wissens auf absehbare Zeit immer unvollständig bleiben werden. Speziell beim graphischen Entwurf, der Kartographie und der Computervisualisierung ist das Wissen nicht leicht formalisierbar, da es nur unvollständig strukturiert und teilweise historisch gewachsen ist. Um den GIS-Visualisierungsentwurf optimal zu automatisieren müßte auch beträchtliches Wissen über die GIS-Anwendung und den Visualisierungskontext formalisiert werden.

Es gibt jedoch noch einige weitere Gründe die dafür sprechen, die GIS-Visualisierungskomponente mit Wissen auszustatten und den Visualisierungsentwurf zu automatisieren, auch wenn er sich nur zum Teil formalisieren und implementieren läßt. Neben der bereits mehrfach erwähnten Tatsache, daß der Visualisierungsentwurf wissensintensiv ist, sprechen folgende weitere Gründe dafür [FeMM92, IEEE92]:

- Der Entwurf von Visualisierungen ist arbeits- und zeitintensiv und die Graphiken werden oft dringend benötigt.
- Die bisher existierenden Visualisierungssysteme gehen vom Entwurf durch trial-and-error aus, eine sicherlich wenig befriedigende Lösung.
- Höchste Qualität der Visualisierungen ist nicht immer notwendig. Die Graphiken werden oft nur einmal benötigt, z.B. bei der explorativen Datenanalyse und können auch manuell noch verbessert werden

## 2.7 Der Lösungsweg: eine Umgebung zur Benutzerunterstützung

Aufgrund der geschilderten Schwierigkeiten löst also weder alleine der erste, noch alleine der zweite Ansatz das Problem zufriedenstellend. Eine vollständige Automatisierung des GIS-Visualisierungsentwurfs scheint wegen des Umfangs und der Komplexität des erforderlichen Wissens nicht möglich. Eine fundierte Ausbildung aller potentiellen Anwender von GIS-Visualisierung kann nicht garantiert werden. Diese Nachteile könnten aber in einer Synthese beider Ansätze weitgehend überwunden werden: Graphisches Wissen sollte einerseits formalisiert und in Expertensystemen implementiert werden, um den GIS-Benutzer beim Entwurf von Visualisierungen direkt zu unterstützen und den Entwurf teilweise oder vollständig zu automatisieren. Andererseits sollte das graphische Wissen aber

auch den Anwendern zugänglich sein, um ihnen bei Bedarf die Möglichkeit einer gezielten Weiterbildung auf dem Gebiet des Visualisierungsentwurfs zu geben (eine solche gezielte Form der bedarfsgesteuerten Weiterbildung wird auch als *learning on demand* bezeichnet [FEER96]).

Der mit dieser Arbeit verfolgte Lösungsweg liegt darin, die GIS-Visualisierungskomponente mit einer wissensbasierten Umgebung zur Benutzerunterstützungssystem auszustatten. Diese Umgebung soll vor allem in der Lage sein, ausdrucksfähige und effektive Visualisierungen von Datensätzen automatisch zu entwerfen, um Benutzern, die nicht über ausreichendes graphisches Vorwissen verfügen, eine möglichst fehlerfreie und nützliche Darstellung zu garantieren. Der Benutzer soll aber auch die Möglichkeit erhalten, sich spezifisches Wissen während der Arbeit mit dem System anzueignen. Die systemseitige Unterstützung der Benutzerweiterbildung sollte dabei idealerweise auf mehreren Ebenen ermöglicht werden: von der Erklärung der Visualisierungen, die automatisch entworfen wurden, bis hin zur Einführung in die Prinzipien der Graphischen Semiotik und der Computervisualisierung.

Das Idealbild ist eine intelligente, offene Umgebung, die Anwender bei der Visualisierung in GIS unterstützt:

- „Intelligent“ heißt, sie soll über graphisches Wissen und Entwurfsalgorithmen auf dem Gebiet der GIS-Visualisierung verfügen, und Visualisierungen automatisch entwerfen und auswerten können.
- „Offen“ heißt, daß das System flexibel auf Benutzeranforderungen reagiert, und sich in einer Vielzahl von GIS-Softwaresystemen und -Umgebungen einsetzen läßt, aber auch, daß es sich im Bezug auf unterstützte Visualisierungstechniken und Anwendungsgebiete erweitern und konfigurieren läßt
- „Unterstützen“ bedeutet, daß das System dem Benutzer den Entwurf vollständig abnehmen kann, ihm aber auch freie Hand in den Bereichen läßt, wo er sich gewachsen fühlt. Unterstützung heißt auch, dem Benutzer so weiterzuhelfen, daß er irgendwann in der Lage ist, den Entwurf auch ohne Systemunterstützung korrekt und effektiv zu bewerkstelligen.
- Der Begriff „Umgebung“ schließt neben dem Entwurfs- und Auswertungssystem auch eine graphische Benutzerschnittstelle ein, die eine möglichst intuitive Spezifikation der Visualisierungsanforderung zuläßt, die Benutzer den Entwurfsprozeß nachvollziehen und seine Ergebnisse überblicken läßt und ihnen Möglichkeiten bietet, das dem System bekannte graphische Wissen zu erwerben.

In ihrem weiteren Verlauf beschäftigt sich diese Forschungsarbeit mit verschiedenen offenen Problemen der Realisierung dieser Umgebung. Zentraler Forschungsgegenstand wird dabei der automatisierte Visualisierungsentwurf für GIS sein. Zur Einarbeitung in diese komplexe Thematik und zur Fundierung des eingeschlagenen Lösungsweges wird Kapitel 3 zunächst einen Überblick über die bisherigen Forschungsarbeiten und den Stand der Technik auf dem Gebiet des automatisierten Entwurfs graphischer Darstellungen geben.

Dabei müssen auch grundlegende Ergebnisse diskutiert werden, die nur indirekt das Anwendungsgebiet GIS berühren.

Anschließend wird untersucht, welche Information über Daten und GIS-Umgebung vorhanden sein müssen, damit effektive GIS-Visualisierungen automatisch entworfen werden können. Neben den Datencharakteristika haben in diesem Zusammenhang die Visualisierungszielsetzung, die Benutzer und die Ausgabemedien eine Bedeutung. Modelle zur Charakterisierung für jede dieser Informationsquellen werden aufgestellt und in einem Modell für effektive GIS-Visualisierung vereint. Dieses Modell bildet schließlich den Ausgangspunkt für die Entwicklung einer Methodik des automatisierten GIS-Visualisierungsentwurfs.

# Kapitel 3

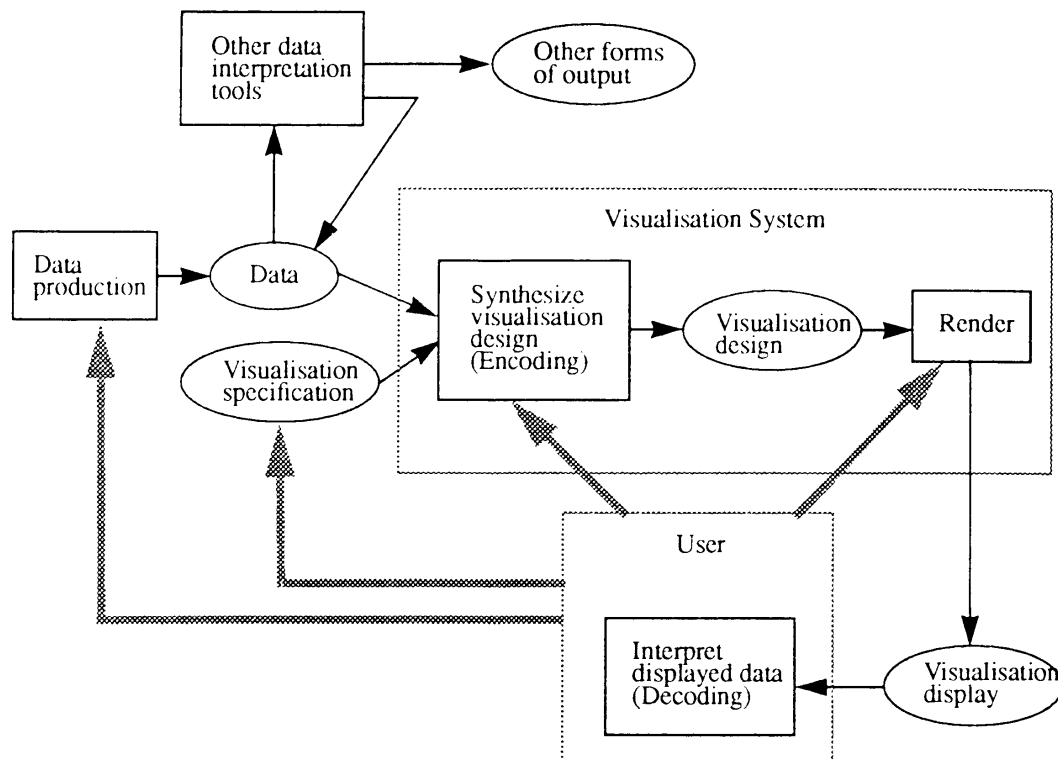
## Automatisierter Visualisierungs- entwurf: Stand der Technik

Das Forschungsgebiet „Automatisierter Entwurf graphischer Darstellungen“ hat in den letzten 10 Jahren ständig an Bedeutung gewonnen. Dies ist auch an der Anzahl der relevanten wissenschaftlichen Veröffentlichungen deutlich zu sehen ist: Sie stieg von ca. 5 pro Jahr Anfang der Achtziger kontinuierlich auf ca. 50 pro Jahr in den Neunzigern an [FeMM92]. Zwei wesentliche Gründe sind für diese Entwicklung verantwortlich. Zum einen wurde nach der stürmischen Entwicklung der Computervisualisierung Ende der achtziger Jahre sehr bald der Bedarf nach einer Systematik für den Visualisierungsentwurf erkannt. Zum anderen legte Mackinlay im Jahr 1986 mit seiner Arbeit zum Entwurf graphischer Darstellungen das Fundament für die Automatisierung des Visualisierungsentwurfs.

In einem wissensbasierten System zur Benutzerunterstützung in der GIS-Visualisierung ist der automatisierte Visualisierungsentwurf eine zentrale Komponente. Deshalb gibt Kapitel 3 dieser Arbeit einen Überblick über den Stand der Technik auf dem Gebiet des automatisierten Entwurfs graphischer Darstellungen. Es werden alle relevanten Vorarbeiten betrachtet und die Ergebnisse anschließend in Kapitel 4 und 5 für den Entwurf von GIS-Visualisierungen verallgemeinert und erweitert.

### 3.1 Charakteristika automatisierter Präsentationssysteme

Eine ganze Reihe automatisierter Präsentationssysteme wurde in den vergangenen 10 Jahren entwickelt [z.B. Mack86b, SeIg91, SeFe91, BeFe94, LSMK95, GuTD97]. Alle diese Systeme wurden für ein spezielles Teilgebiet innerhalb der Computervisualisierung entworfen und alle haben ihre spezifischen Stärken und Schwächen. Robertson und Ferrari [RoFe94] haben die wichtigsten dieser Systeme untersucht und ein Architekturmodell aufgestellt, das die wesentlichen Komponenten solcher automatisierter Präsentationssysteme aufführt: Werkzeuge zur Erzeugung und Vorverarbeitung des Datensatzes, zur Spezifikation der Visualisierung, zur Visualisierungssynthese und für das Rendering (siehe Abb. 3-1).



**Abb. 3-1:** Softwarearchitektur automatisierter Präsentationssysteme [RoFe94]

Um die verschiedenen Systeme direkt vergleichen und bewerten zu können, haben Robertson und Ferrari einen Katalog charakteristischer Attribute aufgestellt. Dieser Katalog eignet sich jedoch nicht nur für den Vergleich bestehender Systeme. Er faßt gleichzeitig auch die wichtigsten Entscheidungen zusammen, die beim Entwurf eines neuen Visualisierungssystems getroffen werden müssen. Die wichtigsten Kriterien sind:

- *Datenmodell:* Welche charakteristischen Merkmale der Datensätze werden in dem System repräsentiert und beim Visualisierungsentwurf beachtet?
- *Visualisierungsspezifikation:* Kann der Anwender seine Wünsche und Zielvorstellungen an die Visualisierung spezifizieren und welche Ausdrucksmittel bietet das System hierfür?
- *Repräsentation der Visualisierungstechniken:* Wie werden die vom System unterstützten Visualisierungstechniken und ihre charakteristischen Merkmale repräsentiert?
- *Matching-Prozeß:* Mit welchen Mitteln wird eine zu Datensatz und Spezifikation passende Visualisierung gefunden?

## 3.2 Automatisierter Visualisierungsentwurf: Zwei Fallstudien

Zwei der wichtigsten und auch für diese Arbeit grundlegenden Systeme sollen im Detail betrachtet und verglichen werden. Das erste System stammt aus der Mitte der achtziger Jahre und wurde schon in der Einleitung dieses Kapitels genannt: es handelt sich um APT (A Presentation Tool), das von Mackinlay für die automatische Präsentation relationaler Information in Form von Diagrammen und Graphen entwickelt wurde. Das zweite System entstand Anfang der neunziger Jahre und ist sehr stark von APT beeinflusst worden: VISTA (Visualization Tool Assistant), ein System für den automatisierten Entwurf technisch-wissenschaftlicher Visualisierungen.

### 3.2.1 APT: Automatisierte Präsentation relationaler Informationen

Mackinlay basierte den automatisierten Entwurf in APT [Mack86a] auf die graphische Semiotik von Jacques Bertin. Mackinlay verwendete die Erkenntnis, daß graphische Darstellungen als Sätze einer „graphischen Sprache“ betrachtet werden können und entwickelte eine Kompositionsalgebra für die Synthese graphischer Entwürfe aus einem Satz primitiver Darstellungstechniken. Ein weiterer wichtiger Beitrag war der systematische Einsatz von Kriterien für die Ausdrucksfähigkeit und Effektivität graphischer Darstellungen in Form von Regeln eines wissensbasierten Systems.

#### Datenmodell

Datensätze in APT unterliegen dem relationalen Datenmodell. Als charakteristische Eigenschaften der Relationen werden die unabhängigen und abhängigen Variablen sowie funktionale Abhängigkeiten zwischen den Variablen repräsentiert. Es wird zwischen nominalen, ordinalen und quantitativen Variablen unterschieden und die Wertemenge jeder Variablen repräsentiert. Bei nominalen und ordinalen Variablen geschieht das durch eine Auflistung, bei quantitativen Variablen durch Angabe von Maximum und Minimum des Wertebereichs.

#### Spezifikation der Visualisierung

Für die Spezifikation der Zielsetzung einer Visualisierung stehen in APT nur sehr beschränkte Ausdrucksmittel zur Verfügung. Der Benutzer gibt die Relationen oder Teilrelationen an, die dargestellt werden sollen. Die Reihenfolge in der diese angegeben werden, impliziert Prioritäten innerhalb des Datensatzes: die zuerst genannte Relation hat höchste Priorität und für sie werden die effektivsten Darstellungstechniken reserviert.

#### Repräsentation der Techniken

Jede Darstellungstechnik wird in APT durch einen Satz von Regeln zu ihrer Ausdrucksfähigkeit repräsentiert. Auch für die verschiedenen Operationen zur Zusammensetzung (Komposition) von Darstellungen liegen Ausdrucksfähigkeitsregeln vor. Die Effektivität der Techniken wird nicht direkt repräsentiert, sondern über den Umweg der visuellen Va-

riablen, die sie einsetzen. Die visuellen Variablen selbst sind nach ihrer Effektivität für die Repräsentation von nominalen, ordinalen und quantitativen Daten geordnet.

### Matching-Prozeß

Der Prozeß des Entwurfs einer Darstellung für einen Datensatz besteht aus drei Stufen: Dekomposition, Matching und Komposition. Zunächst wird der Datensatz solange in Teilrelationen zerlegt, bis alle Teilrelationen durch eine primitive (d.h. nicht zusammengesetzte) Darstellungstechnik ausgedrückt werden können. Den Datensatzteilen wird jeweils die effektivste Darstellungstechnik zugeordnet, unter Berücksichtigung der Prioritäten der Relationen. Die so entstandenen primitiven Darstellungen werden in der Kompositionsstufe nacheinander zusammengesetzt, solange bis eine Gesamtdarstellung des vollständigen Datensatzes entstanden ist.

Der Entwurfsraum wird von APT nach diesen drei Stufen in einem Backtrackingverfahren durchsucht. Dabei wird immer höchstens eine ausdrucksfähige Visualisierung bestimmt: Es ist diejenige, die für den Datensatz am effektivsten ist. Die Effektivitätsordnung der visuellen Variablen sortiert den Entwurfsraum für jeden Datensatz vollständig.

Diese drei Stufen des Entwurfs in APT sollen im folgenden detailliert betrachtet werden.

### Dekomposition von Relationen

APT zerlegt Relationen in einem rekursiven Prozeß in Teilrelationen, die die Ausdruckskriterien von primitiven Visualisierungstechniken erfüllen. Dabei verwendet es drei verschiedene Zerlegungsoperationen für Relationen:

- *Funktionale Dekomposition:* Sie spaltet die Menge der abhängigen Variablen einer Relation auf. Aus einer Relation  $R : A \rightarrow B, C$  werden die zwei Teilrelationen  $R_1 : A \rightarrow B$  und  $R_2 : A \rightarrow C$ .
- *Definitionsbereich-Dekomposition:* Hier wird eine unabhängige Variable  $U$  der Relation ausgewählt und die Relation durch  $|U|$  Teilrelationen ersetzt. Jede dieser neuen Relationen wird gebildet, indem man aus der ursprünglichen Relation alle Tupel auswählt, die einen bestimmten, festen Wert für  $U$  haben und anschließend  $U$  aus diesen Tupeln entfernt. So entstehen z.B. aus  $R : A, B \rightarrow C$  mit  $B = \{a, b, c\}$  die drei Relationen  $R_a : A \rightarrow C$ ,  $R_b : A \rightarrow C$  und  $R_c : A \rightarrow C$ .
- *Bildbereich-Dekomposition:* Sie entfernt eine abhängige Variable aus einer Relation. Aus  $R : A \rightarrow B, C$  wird  $R' : A \rightarrow B$ . Die Variable  $C$  wird bei der später folgenden Komposition wieder in  $R'$  eingesetzt.

Die drei Dekompositionsoperationen sind so gewählt, daß zu ihnen analoge Kompositionsoperationen für Visualisierungen definiert werden können. Visualisierungen der entstehenden Teilrelationen werden dann in einer späteren Phase durch eine analoge Kompositionsoperation zusammengesetzt. Die entstehende zusammengesetzte Visualisierung ist ausdrucksfähig für die ursprüngliche Relation.

Die Dekompositionsoperationen haben, mit geeigneten Ergänzungen, außerdem die Eigenschaft, daß sie informationserhaltend sind. Eine Dekomposition einer Relation ist in-



formationserhaltend, wenn die Summe der in den Teilrelationen enthaltenen Information gleich der in der ursprünglichen Relation enthaltenen Information ist. Dies ist ganz offensichtlich eine Voraussetzung für den Entwurf ausdrucksfähiger Visualisierungen.

### Ausdrucksfähigkeit und Komposition

Jeder Darstellungstechnik ist in APT eine Regel zur Ausdrucksfähigkeit zugeordnet. Solche Regeln prüfen für einen gegebenen Datensatz, ob er sich mit einer Technik darstellen läßt. Diese Prüfung betrifft vor allem die Struktur der Relation und die Ordnungstypen der enthaltenen Variablen. Für Balkendiagramme besagt die Ausdrucksfähigkeitsregel beispielsweise, daß der Datensatz eine Funktion  $f : A \rightarrow B$  sein muß, die unabhängige Variable  $A$  nicht quantitativ und die abhängige Variable  $B$  nicht nominal sein darf. Damit die Balkendiagramme nicht zu breit werden, wird zusätzlich  $|A| < 25$  gefordert.

Ein typisches Balkendiagramm, wie es von APT nach dieser Regel erzeugt wurde, ist in Abb. 3-2 zu sehen. Die unabhängige Variable  $A$  enthält Automarken und die abhängige Variable  $B$  die entsprechenden Preise für das Jahr 1979. Diese Relation erfüllt offensichtlich die Ausdruckskriterien für Balkendiagramme.

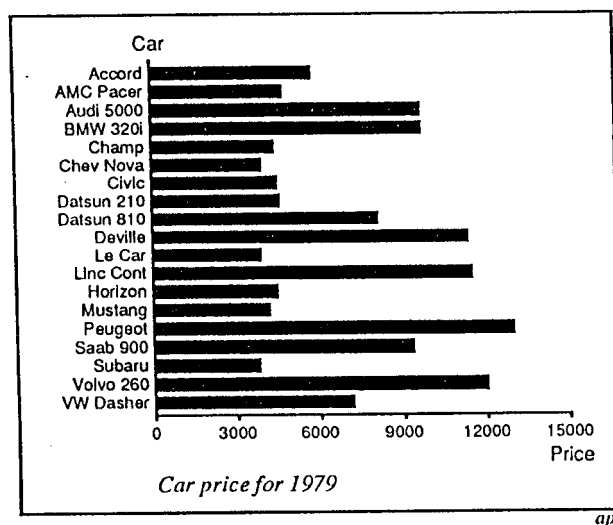


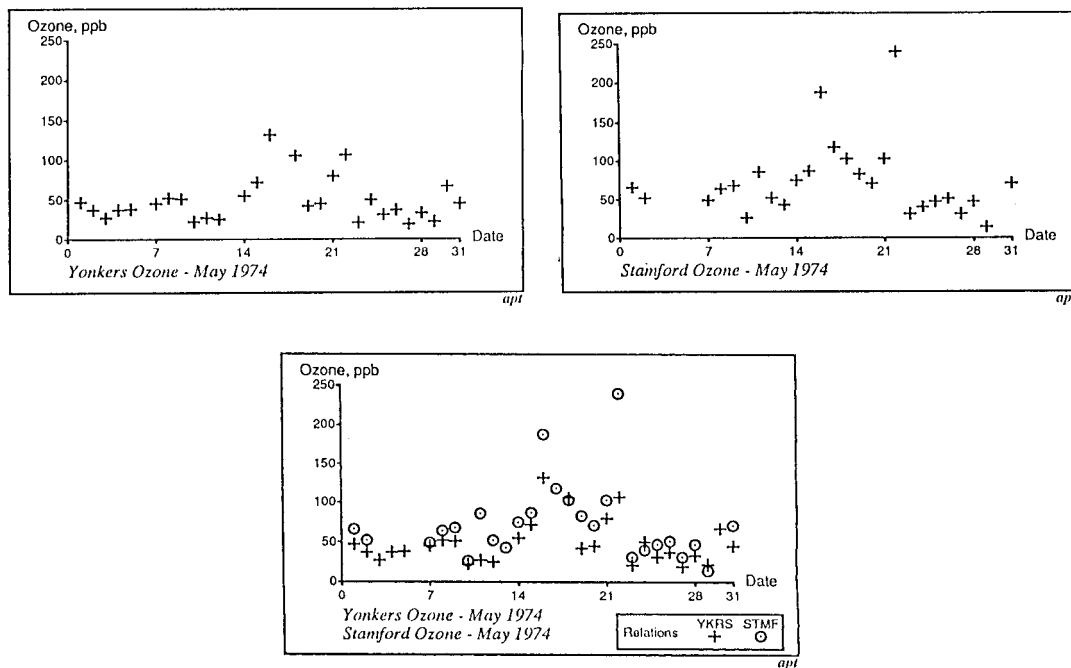
Abb. 3-2: Automatische Präsentation mit APT: ein Balkendiagramm [Mack86a]

Die Ausdrucksregeln in APT erzeugen auch gleichzeitig eine symbolische Beschreibung der Darstellung und weisen ihr die benötigten visuellen Variablen zu. In jedem Schritt wird dabei immer unter allen passenden visuellen Variablen die effektivste zugewiesen. Da priorisierte Relationen zuerst bearbeitet werden, erhalten sie automatisch die effektivste Darstellungsart.

### Komposition

Eine weitere Besonderheit, die APT vor seinen Vorläufersystemen auszeichnet, ist eine Kompositionsalgebra für visuelle Darstellungen. Kompositionsoperatoren setzen zwei vi-

suelle Darstellungen zusammen, indem diejenigen Teile verschmolzen werden, die die gleiche Information ausdrücken. Eine solche Operation ist die Zweiachsenkomposition. Die zwei oberen Diagramme in Abb. 3-3 haben identische Achsen, nämlich Ozonkonzentrationen auf der vertikalen und Tage auf der horizontalen Achse. Beide Achsen der zwei Diagramme lassen sich deswegen verschmelzen und ein zusammengesetztes Diagramm entsteht. Zur Unterscheidung werden im zusammengesetzten Diagramm automatisch zwei unterschiedliche Signaturen verwendet.



**Abb. 3-3:** Zweiachsenkomposition: aus den beiden oberen Diagramme entsteht das untere [Mack86]

APT unterstützt noch zwei weitere Kompositionsoperatoren:

- *Einachsenkomposition:* Mit dieser Operation werden Darstellungen zusammengesetzt, die nur in einer Achse übereinstimmen.
- *Signaturkomposition:* Mit dieser Operation werden die Signaturen zweier Darstellungen verschmolzen. Dazu muß zu jeder Signatur in der einen Darstellung eine korrespondierende in der anderen Darstellung existieren und die Signaturen müssen unterschiedlichen visuellen Variablen zugehören.

## Diskussion

Die wichtigsten neuen Aspekte von APT wurden schon in der Einleitung zu diesem Kapitel genannt. Es waren die Formalisierung und softwaretechnische Realisierung von Ausdrucksfähigkeit, Effektivität und der Komposition graphischer Darstellungen.

Das System weist allerdings auch einige gravierende Lücken auf. Insbesondere kann es keine Information über die Zielsetzung der Darstellung repräsentieren und erzeugt deswegen für jeden Datensatz nur genau eine Darstellung. Auch können nur sehr einfache Charakteristika des Datensatzes ausgedrückt und im Entwurf berücksichtigt werden. Der Satz an Darstellungstechniken, die das System unterstützt ist sehr beschränkt: es sind lediglich einige Formen von Diagrammen und Graphen.

### 3.2.2 VISTA: Wissensbasiertes System für den Visualisierungsentwurf

VISTA ist ein System für den automatisierten Entwurf technisch-wissenschaftlicher Visualisierungen. Es entstand Anfang der neunziger Jahre und baut auf APT und der Arbeit von Mackinlay auf. VISTA unterstützt jedoch im Vergleich zu den einfachen Diagrammdarstellungen von APT wesentlich mehr und komplexere Visualisierungstechniken.

#### Datenmodell

Das Datenmodell von VISTA ist wesentlich vollständiger, wie Abb. 3-4 zeigt. Relationen und ihre qualitativen (d.h. nominalen) und quantitativen Variablen werden jeweils durch spezifische Attributlisten beschrieben. Attribute der Relationen sind u.a. Abhängigkeiten zwischen Variablen, das verwendete Koordinatensystem und der Abstand der Meßpunkte. Für quantitative Variablen sind dies Minimum, Maximum, Kardinalität, Maßeinheit, etc.

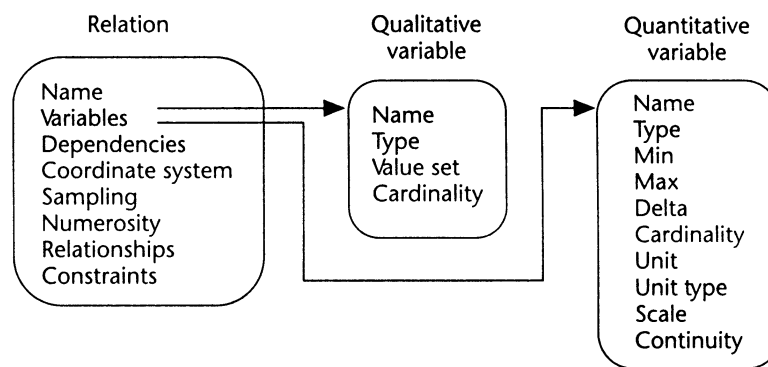


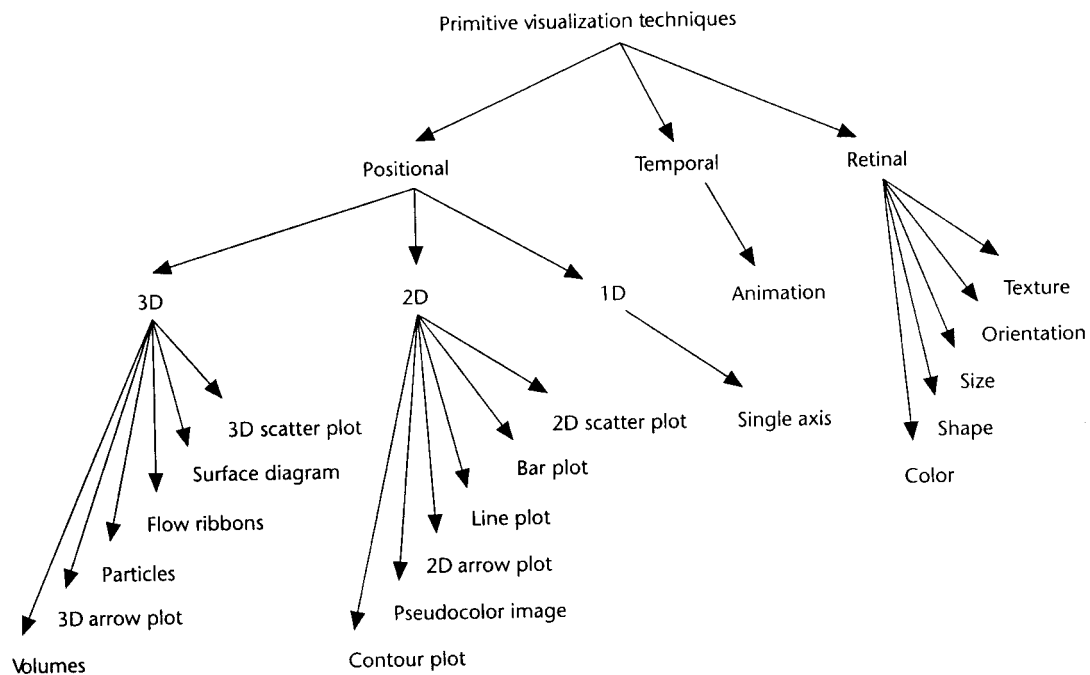
Abb. 3-4: Datencharakterisierung in VISTA [SeIg94]

#### Spezifikation der Visualisierung

Die Möglichkeiten zur Spezifikation der Visualisierung sind auch in VISTA sehr bescheiden. Visualisierungsziele können ebenso wenig ausgedrückt werden wie Priorisierungen einzelner Relationen.

#### Repräsentation der Techniken

Abb. 3-5 zeigt alle in VISTA unterstützten „primitiven“ Visualisierungstechniken. Die Techniken werden in VISTA ebenfalls durch Regeln zur Ausdrucksfähigkeit repräsen-



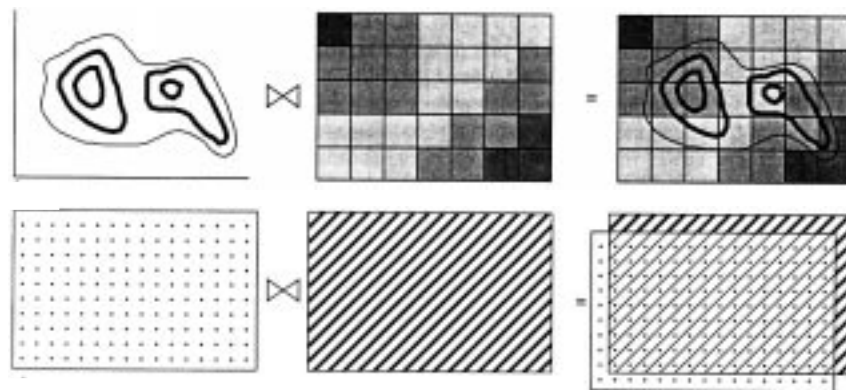
**Abb. 3-5:** Primitive Visualisierungstechniken in VISTA [SeIg94]

tiert. Zusätzlich gibt es aber auch explizite Regeln zur Effektivität der Techniken. Senay und Ignatius haben die Regeln aus der Literatur zu den Visualisierungstechniken extrahiert und formalisiert [SeIg90b].

### Matching-Prozeß

Der Visualisierungsentwurf erfolgt, ähnlich wie bei APT, in drei Stufen. Die Datensätze werden zunächst solange zerlegt, bis Teilrelationen mit einer abhängigen und bis zu vier unabhängigen Variablen vorliegen. Mit den Ausdrucksregeln werden zu allen Teilrelationen passende Visualisierungstechniken bestimmt und anschließend in einer Kompositionsphase die Teilvisualisierungen zusammengesetzt.

VISTA verallgemeinert und ergänzt die Kompositionsooperatoren von APT, so daß sie auch für dreidimensionale Visualisierungen gelten und die Fähigkeiten moderner Graphikworkstations und Renderingsysteme ausnutzen. Abb. 3-6 zeigt zwei der spezifischen Kompositionen in VISTA. Der erste Operator überlagert eine flächenförmige Visualisierung mit einer linienförmigen. Der zweite Operator überlagert zwei flächenförmige Visualisierungen und stellt eine von ihnen transparent dar.



**Abb. 3-6:** Visualisierungskomposition durch Überlagerung [SeIg94]

### 3.3 Charakterisierungsschemas für den Visualisierungsentwurf

In diesem Unterkapitel werden einige theoretische Arbeiten im Zusammenhang mit dem automatisierten Visualisierungsentwurf vorgestellt. Es handelt sich dabei in erster Linie um Charakterisierungsschemas für drei wichtige Einflußfaktoren des Visualisierungsentwurfs: Datensatz, Zielsetzung und Kontext der Visualisierung.

Unter dem Begriff *Visualisierungskontext* werden diejenigen Randbedingungen einer Visualisierung zusammengefaßt, die nicht direkt den Datensatz und die Visualisierungszielsetzung betreffen. Je nach Anwendungsgebiet und Funktionalität der Visualisierungssoftware können sehr unterschiedliche Aspekte darunter verstanden werden. Feiner, Mackinlay und Marks [FeMM92] haben folgende Elemente einer Beschreibung des Visualisierungskontexts vorgeschlagen:

- *task domain knowledge*: Allgemeines Wissen über das Anwendungsgebiet.
- *communicative goals*: Was soll mit der Visualisierung mitgeteilt werden?
- *situation model*: Die spezifische Situation, für die die Visualisierung erzeugt werden soll, z.B. Notfall- im Gegensatz zu Routinesituationen.
- *user model*: Fähigkeiten und Ausbildung des Benutzers.
- *resource model*: Charakterisierung der Rendering-Software und -Hardware, der Ausgabe- und Interaktionsgeräte, etc.

Von diesen fünf Kontextaspekten sind im Zusammenhang mit dem Visualisierungsentwurf in der wissenschaftlichen Literatur bisher nur die Benutzermodelle ausführlicher untersucht worden. Benutzermodelle bilden deswegen neben den Datencharakteristika und der Visualisierungszielsetzung das Thema dieses Kapitels über Charakterisierungsschemas.

### 3.3.1 Datencharakteristika

#### Roth und Mattis

Ein erster Ansatz zur Beschreibung der Datencharakteristika wurde bereits mit den Datenmodell von VISTA vorgestellt. Etwas früher haben schon Roth und Mattis eine detaillierte Charakterisierung entwickelt [RoMa90]. Sie war für die Visualisierung relationaler Information in Form von Diagrammen mit dem System SAGE geplant. Die Charakterisierung umfaßt einige Merkmale, die auch in VISTA beschrieben werden, ermöglicht aber weitergehende Angaben über die Semantik des Datensatzes:

- *Koordinaten vs. Anzahlen:* Sind die Datenwerte Koordinaten, wie z.B. Datumsangaben, oder Anzahlen, wie z.B. Preise? Koordinaten sollten mit einem Scatter-Plot, Anzahlen dagegen mit einem Balkendiagramm dargestellt werden.
- *Strukturelle Eigenschaften:* Zum Beispiel: zwei Variablen einer Relation sind die Ober- und Untergrenze eines Intervalls. Das Intervall sollte in Form eines Balken dargestellt werden.
- *Algebraische Abhängigkeiten:* Zum Beispiel: zwei Variablen einer Relation sind Mächtigkeiten von Teilen eines Ganzen. Sie können also summiert werden, wie in:  $\text{Männer} + \text{Frauen} = \text{Gesamtbevölkerung}$ .

#### Brodlie et. al.

Sehr nützlich und kompakt ist auch die Notation von Brodlie et. al. [Brod92], die in erster Linie entwickelt wurde, um Visualisierungsmethoden zu klassifizieren. Sie bezeichnet die Struktur der Entities (Datenobjekte), die eine Visualisierungsmethode ausdrücken kann. Diese Notation hat die Form

$$E_D^B,$$

wobei  $D$  den Definitionsbereich und  $B$  den Bildbereich bezeichnet.

Beim Definitionsbereich werden folgende drei Fälle unterschieden:

- Punkte aus einer kontinuierlichen Menge (z.B. die Punkte der Ebene:  $E_2$ )
- diskrete Bereiche aus einer kontinuierlichen Menge (z.B. statistische Bezirke in einer Karte:  $E_{[2]}$ )
- Elemente aus einer Aufzählung (z.B. Landnutzungsarten:  $E_{\{1\}}$ )

Beim Bildbereich sind die wichtigsten Fälle:

- Punkte:  $E^P$
- Skalare:  $E^S$
- Vektoren, z.B.  $E^{V_2}$  für 2D-Vektoren

In dieser Notation ist beispielsweise eine Choroplethenkarte vom Typ  $E_{[2]}^S$ , ein 3D-Vektorfeld ist vom Typ  $E_3^{V_3}$ .

### 3.3.2 Visualisierungszielsetzung

#### Visualisierungsaufgabe nach Roth und Mattis

Roth und Mattis haben neben der Datencharakterisierung auch eine Liste von Aufgaben erstellt, die ein Anwender mit Hilfe einer Visualisierung ausführen möchte. Sie unterscheiden folgende fünf Aufgaben [RoMa90]:

- Ablesen exakter Werte
- Vergleich von Werten innerhalb einer Relation
- paarweiser Vergleich von Werten zweier verschiedener Relationen
- Erkennen der Werteverteilung einer Variablen
- Erkennen einer Korrelation zwischen Variablen

#### Wehrends Operationsklassen

Etwa zur gleichen Zeit hat sich Stephen Wehrend [WeLe90] mit der Klassifizierung von Visualisierungstechniken beschäftigt. Aus der Visualisierungsliteratur hatte Wehrend Referenzen auf etwa 400 verschiedene Techniken zusammengestellt und versucht, diese verschiedenen Visualisierungsproblemen zuzuordnen. Visualisierungsprobleme wurden durch eine Matrix mit zwei Dimensionen beschrieben: Objekte und Operationen. Unter Objekten werden Elemente des Datensatzes verstanden und es werden unter anderem folgende Objekte unterschieden: Skalare, Skalarfelder, Vektoren, Vektorfelder, Positionen und Geo-Objekte. Auf diesen Objekten definiert Wehrend nun verschiedene visuelle Analyseoperationen, welche durch die Visualisierung unterstützt werden sollen.

Wehrends Klassifizierung von Visualisierungsproblemen ist sehr allgemein und nicht immer eindeutig. Sie ist deswegen eher dafür geeignet, Benutzern von Visualisierungssystemen das Auffinden geeigneter Techniken zu erleichtern als in einem automatisierten System realisiert zu werden. Trotzdem sollen hier einige der Operationen auf Objekten genannt werden, da sie Aspekte möglicher Visualisierungszielsetzungen aufzeigen:

- *Identifizieren:* Alle Objekte sollen gut identifiziert und unterschieden werden können.
- *Auffinden:* Spezifische Objekte sollen in der Visualisierung hervorgehoben und leicht aufgefunden werden können.
- *Unterscheiden:* Objekte sollen leicht unterscheidbar sein.

- *Cluster*: Es soll herausgefunden werden, ob Objekte geclustert auftreten oder nicht.
- *Verteilung*: Die Verteilung von Objekten soll gezeigt werden.
- *Vergleich*: Der Vergleich von Objekten soll ermöglicht werden.
- *Korrelation*: Zwischen verschiedenen Objekten soll eine mögliche Korrelation erkannt werden.

### Robertsons Matrix der Interessenattribute

In seiner Arbeit [Robe91] zum *Natural Scene Paradigm* bei der Visualisierung hat P. K. Robertson auch eine Klassifizierung von Benutzerinteressen an Visualisierungen aufgestellt. Er unterscheidet zunächst zwischen drei möglichen Interessenattributen, die ein Benutzer bezüglich einer Datenvariablen artikulieren kann:

- *Punkt*: Es interessieren lediglich die Werte an verschiedenen Punkten.
- *lokal*: Es interessiert die örtliche Werteverteilung (lokale Gradienten und Muster).
- *global*: Es interessiert die globale Werteverteilung (Trends und Strukturen).

Diese Attribute korrespondieren jeweils mit der elementaren, mittleren und höchsten Stufe der Information, wie sie von Jacques Bertin [Bert82] beschrieben wurden.

Das Interesse an der Korrelation zweier Variablen kann in diesem Modell differenziert beschrieben werden. Es wird notiert als  $A_1 : A_2$  mit  $A_1, A_2 \in \{\textit{Punkt}, \textit{lokal}, \textit{global}\}$ . *Punkt : lokal* liest sich beispielsweise als Interesse an „Werten an Punkten der ersten Variablen in Beziehung zur lokalen Variation der zweiten Variablen“. Ein Anwender könnte mit diesen Interessenattributen zum Ausdruck bringen, daß er das gemeinsame Auftreten von besonderen Werten der einen Variablen und von steilen Gradienten oder spezifischen lokalen Verteilungen der anderen Variablen untersuchen will.

Die Interessenattribute an den verschiedenen Variablen eines Datensatzes werden in einer Matrix zusammengestellt. Tabelle 3-1 zeigt das Beispiel einer solchen Matrix für drei Variablen. In den Diagonalfeldern stehen die Interessenattribute bezüglich jeder der Variablen, wobei auch mehrere Attribute gleichzeitig erscheinen dürfen. In den Feldern oberhalb der Diagonale stehen für jedes Variablenpaar die Interessenattribute an der Korrelati-



on dieser beiden Variablen. Auch hier können mehrere Interessen gleichzeitig verfolgt werden.

	Variable 1	Variable 2	Variable 3
Variable 1	<i>Punkt</i>	<i>lokal : Punkt</i>	<i>global : global</i>
Variable 2		<i>Punkt</i> <i>global</i>	<i>point : global</i>
Variable 3			<i>lokal</i> <i>global</i>

**Tabelle 3-1:** Eine Matrix der Interessenattribute, nach [Robe91]

Insgesamt läßt sich über diese Methode der Charakterisierung der Visualisierungszielsetzung sagen, daß sie sehr detaillierte und differenzierte Aussagen ermöglicht. Allerdings werden nicht alle Anwender von Visualisierungssystemen diese Notation verständlich finden und ihre Visualisierungszielsetzung in ihr formulieren können. Viele Anwender werden auch gar nicht eine solch detaillierte Aussage über ihre Zielsetzung machen können. Sehr hilfreich kann die Charakterisierung jedoch in der Datenvisualisierung für das Überprüfen von Hypothesen sein, da diese sehr präzise und detaillierte Fragestellungen an den Datensatz mit sich bringen.

#### Casners logische Taskbeschreibung

Eine weitere Möglichkeit der Beschreibung einer Visualisierungszielsetzung bedient sich der Methoden der Aufgabenanalyse (*Task Analysis*). Stephen M. Casner hat sich mit der Automatisierung des Entwurfs graphischer Darstellungen für ein klar beschränktes Anwendungsfeld beschäftigt. Sein System BOZ entwirft Diagramme und Darstellungen, die der Reservierungsagentur einer Fluggesellschaft helfen, geeignete Flüge für Kunden zu finden [Casn91]. Die Kriterien nach denen Flüge gesucht werden sollen sind klar definiert, wie z.B. günstiger Preis und direkte Verbindung.

Casner definiert eine Sprache zur logischen Taskbeschreibung, in der die Arbeitsschritte einiger typischer Aufgaben der Reservierungsagentur spezifiziert werden können. Eine Aufgabe wird als LISP-Prozedur notiert, ergänzt durch Charakterisierungen der Eingabedaten und durch sogenannte logische Operatoren (LOPs). Die LISP-Prozedur beschreibt detailliert alle Schritte zur Erledigung der Aufgabe, wobei LOPs aufgerufen werden können. Die LOPs sind hauptsächlich Operationen zur Werteabfrage und -interpretation, z.B. „Dauer des Transitaufenthaltes bestimmen“. Das wissensbasierte System BOZ analysiert nun die logische Taskbeschreibung und erstellt eine graphische Darstellung, die den Task unterstützt.

LOPs erlauben eine sehr detaillierte Beschreibung der Zielsetzung einer Visualisierung, können aber nur bei klar definierbaren Abläufen eingesetzt werden. Mit LOPs können Visualisierungen auch von automatisierten Systemen sehr zielgerichtet entworfen werden. In vergleichbaren Anwendungsbereichen, wo gewisse Arbeitsabläufe häufig wiederkeh-

ren und sich ähnlich gut formalisieren lassen wie der Prozeß der Flugreservierung, wird man deswegen sehr gute Resultate mit dem Ansatz von Casner erzielen.

### 3.3.3 Benutzermodell

Benutzermodelle für den automatisierten Visualisierungsentwurf sind bisher weniger gut untersucht worden als Datencharakteristika und Visualisierungszielsetzung. Keines der in der wissenschaftlichen Literatur beschriebenen Prototypsysteme verfügt über ein explizites Benutzermodell. Es gibt aber einige Vorarbeiten, die verschiedene Aspekte der Benutzermodellierung untersuchen, die auch für den GIS-Visualisierungsentwurf von Interesse sind. Zum einen ist hier die Arbeit von Domik und Gutkauf zur Modellierung der individuellen visuellen Wahrnehmung und Kognition von Benutzern zu nennen. Andererseits gab es in der GIS-Forschung in den vergangenen Jahren mehrfach Anstrengungen, Benutzermodelle zielgerichtet einzusetzen. Auch aus diesem Bereich sollen zwei Arbeiten hier exemplarisch vorgestellt werden.

#### Modellierung der visuellen Wahrnehmung und Kognition

Domik und Gutkauf haben die Modellierung der individuellen visuellen Wahrnehmung und Kognition von Benutzern untersucht [DoGu94]. Mit diesen Benutzermodellen sollen insbesondere adaptive Visualisierungssysteme unterstützt werden. Das Modell umfaßt fünf Parameter, die über die individuelle visuelle Wahrnehmung und die Fähigkeiten der räumlichen Kognition und Motorik von Benutzern Auskunft geben:

- *Farbwahrnehmung:* Wie gut und wie schnell werden verschiedene Farben unterschieden? Dieser Parameter fragt insbesondere auch nach Farbenblindheit und gibt Auskunft darüber, welche Farbkodierung in einer Visualisierung für den Benutzer die geeignetste ist.
- *Farbordnung:* Wie schnell können verschiedene Farben in eine „natürliche“ Ordnung gebracht werden, und wie sieht diese Ordnung aus? Benutzt eine Visualisierung eine Farbskala, sollte sie der individuell-intuitiven Farbordnung des Benutzers nicht widersprechen.
- *Farbgedächtnis:* Wie exakt und wie schnell können Farben aus dem Gedächtnis abgerufen werden? Ein gutes Farbgedächtnis ist Voraussetzung dafür, daß Benutzer zwei verschiedene Visualisierungen mit der gleichen Farbsymbolik vergleichen können.
- *räumliches Vorstellungsvermögen:* Wie gut können räumliche Objekte in der Vorstellung rotiert werden? Räumliches Vorstellungsvermögen ist Voraussetzung für die Interpretation aller dreidimensionaler Darstellungen.
- *mechanische Feinkoordinierung:* Wie gut und wie schnell können feine Handbewegungen ausgeführt und mit der Wahrnehmung koordiniert werden? Nur Benutzer mit einer guten Feinkoordinierung der Handbewegungen können mit einem Eingabegerät wie der Maus effektiv eine Visualisierung steuern.

Eine besondere Eigenschaft dieser fünf Parameter ist, daß sie alle relativ gut getestet werden können. Domik und Gutkauf haben verschiedene bekannte Testverfahren aus der Wahrnehmungspsychologie bearbeitet und in Form von Computerspielen implementiert. Das Ergebnis war ein interaktives Testprogramm für alle fünf Parameter, das ein Benutzer zu Beginn seiner Arbeit in einer Visualisierungsumgebung absolvieren kann.

### **Informationstheorie und Benutzermodelle**

Einen ganz anderen Aspekt der Benutzermodellierung untersuchen Lindholm und Sarjakowski [LiSa92]. Sie wenden die Shannon'sche Informationstheorie auf das Problem der Query-Formulierung in GIS an. Ziel ist es, den Informationsgehalt des Query-Resultats in jedem Schritt zu optimieren. Dazu beschreiben sie den Benutzer, von einem pragmatischen, informationstheoretischen Standpunkt ausgehend, mit drei Parametern:

- die Information, die der Benutzer gewinnen möchte (z.B. durch Spezifikation bezüglich Raum, Zeit und Thema)
- die Information, die der Benutzer schon hat
- die Fähigkeit des Benutzers, tiefergehende Information aus einem Stück oberflächlicher Information herzuleiten

Diese Parameter beschreiben Benutzercharakteristika, die eine wesentlich kürzere Geltungsdauer haben als das Modell von Domik und Gutkauf. Sie geben eine andere Sicht des Benutzers wieder, sind aber ebenfalls von Bedeutung für den Visualisierungsentwurf, denn Visualisierungen stellen letztlich für den Benutzer vor allem eine effektive Informationsquelle dar.

## **3.4 Automatisierungsansätze in der computergestützten Kartographie**

Die Verfügbarkeit von Karten in den verschiedensten Maßstäben und Detaillierungsgraden hat in unserer Gesellschaft eine große Bedeutung. Sie ist aber auch sehr kostspielig, da viele der anfallenden Arbeitsschritte von der Datenerfassung bis hin zur Kartenproduktion arbeitsaufwendig sind. Folgerichtig wurde schon seit einiger Zeit versucht, sowohl die Geodatenerfassung, als auch den kartographischen Entwurfsprozeß zu automatisieren.

In der kartographischen Forschung spielte die Automatisierung der Generalisierung, einer Stufe des kartographischen Entwurfs, in den letzten 10 - 15 Jahren eine wichtige Rolle. Untersucht wurde vor allem die Generalisierung von topographischen Karten aus Geodatenbanken. Mit der „quantitativen Revolution“ in der Geographie trat aber seit Mitte der achtziger Jahre auch der automatisierte Entwurf thematischer Karten immer mehr in den Vordergrund [MuJV86]. Insgesamt sind die praktischen Resultate aus langen Jahren der Generalisierungsforschung relativ bescheiden. Topographische Karten werden immer noch „per Hand“ aus größeren Maßstäben generalisiert und nur Teilaufgaben, wie die Beschriftung, wurden bisher erfolgreich automatisiert.

Um den Rahmen dieser Arbeit nicht zu sprengen, können aus der umfangreiche Literatur zur Automatisierung der Kartographie hier nur einige neuere Arbeiten exemplarisch vorgestellt werden. Der Schwerpunkt liegt naturgemäß auf den Arbeiten zur kartographischen Symbolisierung und zum Entwurf thematischer Karten, da sie für die GIS-Visualisierung von besonderer Bedeutung sind.

### 3.4.1 Der kartographische Entwurfsprozeß

Einen guten Überblick über den kartographischen Entwurfsprozeß und die verschiedenen Ansätze, ihn zu automatisieren gibt ein Artikel von Buttenfield und Marks [BuMa91a]. Er folgt der traditionellen Gliederung des Entwurfsprozesses in drei Stufen: Generalisierung, Symbolisierung und Produktion.

#### Kartographische Generalisierung

Generalisierung ist die erste Stufe des kartographischen Entwurfs. Sie umfaßt die Modellierung geographischer Phänomene und die Anpassung des entstandenen Modells an die gewünschte Sicht. Auf einer Karte können nie alle Phänomene des korrespondierenden Weltausschnitts gleichzeitig und vollständig dargestellt werden. Es ist also zunächst erforderlich auszuwählen, welche Information auf der Karte dargestellt und welche Information weggelassen wird. Diese Auswahl wird sich u.a. danach richten, welches Thema auf der Karte porträtiert wird und welchen Maßstab sie erhalten soll. In zweiter Instanz ist häufig eine Abstraktion der ausgewählten Information nötig, damit sie mit den beschränkten Mitteln des zweidimensionalen graphischen Systems dargestellt werden kann. Die Generalisierung läßt sich insgesamt in drei Unterpunkte aufgliedern:

- Die *Vereinfachung* umfaßt die Reduktion der Repräsentationsgenauigkeit der Kartenelemente, die Auswahl relevanter Elemente und Elementbeziehungen und die Verschiebung von Elementen zur Eliminierung unwichtiger Details.
- Die *Klassifikation* umfaßt die Aggregation von benachbarten Elementen zu Einheiten und die Partitionierung metrischer Größen in einige Klassen.
- Die *Verbesserung* umfaßt das gezielte Einfügen neuer Information in die Karte zur Vereinfachung der Interpretation, wie beispielsweise Interpolation und Rekonstruktion von Elementen.

Abb. 3-7 stellt wichtige Operationen der Generalisierung anschaulich dar.

#### Kartographische Symbolisierung

Mit der Symbolisierung, dem zweiten Schritt des kartographischen Entwurfs, wird die im ersten Schritt ausgewählte und aufbereitete Information in graphische Zeichen umgesetzt. Dabei werden Signaturen, Linienstärken, Farben, etc. zugewiesen. Hier muß sichergestellt werden, daß die Signaturen jeweils zur Darstellung der Information geeignet sind (Ausdrucksfähigkeit) und daß sie für den Betrachter einfach und intuitiv zu interpretieren sind (Effektivität). Wie diese Forderung im Einzelfall realisiert werden kann, darüber gibt z.B. die graphische Semiotik Bertins Auskunft, die schon in Abschnitt 2.5.2 besprochen




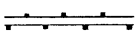
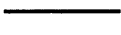

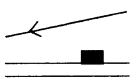








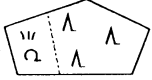





Elementarer Vorgang	Darstellung in der		
	Ausgangskarte	neuen Karte	
	Maßstab der Ausgangskarte		neuen Karte
Rein geometrische Generalisierung			
1. Vereinfachen			
2. Vergrößern (vor allem Verbreitern)			
3. Verdrängen (Folge von 2)			
Geometrisch-begriffliche Generalisierung			
4. Zusammen- fassen			
5. Auswählen (bzw. Fortlassen)			
6. Klassifizieren bzw. Typisieren (einschließlich Umwandeln in Signaturen)			
7. Bewerten (z. B. Betonen)			

Abb. 3-7: Kartographische Generalisierung [Hake82]

wurde. Zu den allgemeingültigen Regeln der Semiotik können aber, je nach dem beabsichtigten Einsatz der Karte, noch andere, situations- und anwendungsspezifische Bedingungen hinzukommen. Wird die Karte beispielsweise auf einem Computerdisplay angezeigt, ist die Auflösung sehr viel niedriger als bei Papierkarten und es sind andere Signaturen effektiver.

### Kartenproduktion

Die letzte Stufe des Entwurfsprozesses besteht darin, aus der Sammlung graphischer Signaturen eine geschlossene graphische Darstellung zu erzeugen. Hier müssen wieder die Restriktionen des zweidimensionalen graphischen Systems „Karte“ beachtet werden. Operationen der Kartenproduktion sind:

- Das *Layouten*, dessen Ziel eine übersichtliche und konsistente Gesamtgestaltung der Karte ist,
- die *Verdrängung* von überlappenden Signaturen auf der Karte (gelegentlich auch schon in der Generalisierungsstufe behandelt),
- und die *Beschriftung* der Kartenelemente.

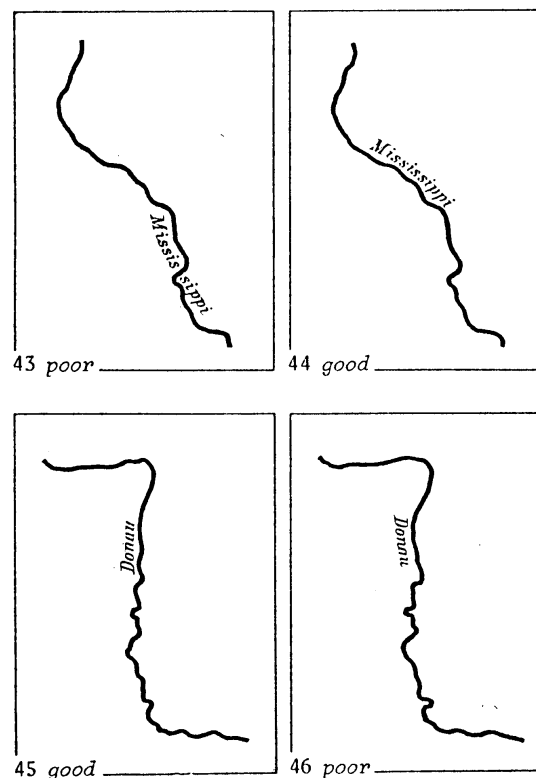
### 3.4.2 Automatisierte Kartenbeschriftung

Dieser und die folgenden beiden Abschnitte stellen einige Arbeiten und Prototypsysteme vor, die Teilaspekte dieses Entwurfsprozesses automatisieren. Der Schwerpunkt liegt bei dieser Diskussion auf der Symbolisierung, die für den GIS-Visualisierungsentwurf am wichtigsten ist. Auf die Kartenproduktion wird nur exemplarisch, anhand eines Systems zur automatischen Kartenbeschriftung eingegangen.

Die Beschriftung von Karten ist ein Problem, dessen Feinheiten sich nur schwer formal ausdrücken lassen, das jedoch gut automatisiert werden kann, wenn man gewisse Einschränkungen in Kauf nimmt [Free91]. In einem berühmten Artikel hat der Schweizer Kartograph Eduard Imhof versucht, die ästhetischen und semantischen Aspekte der Kartenbeschriftung auszudrücken ([Imho75], auf Deutsch schon 1962 erschienen). Dies ist ihm in vorbildlicher Weise gelungen, es waren aber in diesem Artikel von etwas über 20 Seiten mehr als 100 Abbildungen, Beispiele und Gegenbeispiele nötig, um die Feinheiten dieses Themas zufriedenstellend zu dokumentieren (siehe Abb. 3-8).

Im Gegensatz dazu konnte Jones, der sich allerdings auf das einfachere Problem der Beschriftung von Punkten beschränkte, mit einem einfachen Prolog-Programm bereits gute Resultate erzielen [Jones89]. Sein Programm ließ nur sechs mögliche Positionen der Beschriftung von Punkten, z.B. Städten zu. Diesen sechs Positionen ordnete er jeweils einen Eignungsgrad zu, den er aus der Arbeit von Imhof ermittelt hatte. Er formulierte die Kartenbeschriftung als Optimierungsproblem, bei dem es gilt, für jeden Punkt die Eignung der Beschriftung zu optimieren. Diese Optimierung konnte er für nicht allzu große Karten mit einem Backtracking-Algorithmus effizient lösen.

Auch die deutsche Forschung hat sich in den vergangenen Jahren mit dem Problem der Kartenbeschriftung beschäftigt und dabei sehr interessante Resultate erzielt. Wolfgang Kresse hat in seiner Dissertation aus dem Jahr 1994 die Kartenbeschriftung sehr viel umfassender als Jones als Optimierungsproblem formuliert und mit einem Verfahren der linearen Programmierung auch für große Karten effizient gelöst [Kres94].



**Abb. 3-8:** Beschriftung von Flüssen [Imho75]

Eine Arbeitsgruppe um Frank Wagner an der Freien Universität Berlin hat unterdessen erstmals nachgewiesen, daß das Problem NP-vollständig ist [WaWo95]. Sie konnten zeigen, daß sich das Beschriftungsproblem auf das bekannte 3-SAT-Problem zurückführen läßt. Zusätzlich haben sie ein heuristisches Lösungsverfahren entwickelt, mit dem sich sehr gute Näherungslösungen erzielen lassen. Das Verfahren führt das Beschriftungsproblem mit gewissen Abstrichen auf 2-SAT zurück, das bekanntermaßen polynomial ist.

### 3.4.3 Architekturen kartographischer Expertensysteme

#### Ein kartographisches Entwurfsmodell für Expertensysteme

Forrest hat den kartographischen Entwurfsprozeß unter dem Gesichtspunkt der Automatisierung untersucht [Forr90]. Dabei leitete er ein kartographisches Entwurfsmodell für Expertensysteme ab, dessen Schwerpunkt bei den thematischen Karten liegt. Das Modell unterscheidet sich von der klassischen Dreiteilung in Generalisierung, Symbolisierung und Produktion dadurch, daß es konkretere Entwurfsschritte spezifiziert und dabei auch Interaktionen mit dem Kartenautor berücksichtigt. Forrest gibt zwar nur sehr allgemeine Realisierungshinweise, sein Modell kann aber als Grundlage für weitere Forschungen hinsichtlich eines thematisch-kartographischen Expertensystems dienen. Die vier wesentlichen Schritte des Modells sind:

1. *Beschreibung:* Der Kartenautor spezifiziert den Informationsgehalt der gewünschten Karte. Dazu zählt nach Forrest auch die Detailstufe, in der Information dargestellt werden soll und der Zweck der Visualisierung. Der Kartenbenutzer, der vom Kartenautor verschieden sein kann, soll ebenfalls charakterisiert werden.
2. *Layout:* In dieser Stufe werden Entscheidungen getroffen, welcher geographische Bereich dargestellt wird, in welchem Maßstab das geschieht und welche Dimensionen die resultierende Karte hat. Die Entscheidungen liegen zum Teil beim Benutzer, teilweise können sie auch schon vom Expertensystem aus anderer Information hergeleitet werden.
3. *Grundkarte:* Das System bestimmt bzw. entwirft die Grundkarte, die als Basis für die thematische Karte dienen soll. Einige Operationen der Generalisierung können an dieser Stelle erforderlich sein, insbesondere wenn der Maßstab der Grundkarte verändert werden muß oder wenn sie zu komplex ist um zusätzliche thematische Information aufzunehmen. (Forrest verwendet für diese Stufe den etwas unklaren Begriff „Datenauswahl“.)
4. *Symbolisierung:* Die ausgewählte Information wird durch kartographische Signaturen dargestellt. Forrest gibt hier eine Klassifizierung thematisch-kartographischer Repräsentationen an.
5. *Darstellung und Modifikation:* Die Karte wird dargestellt. Der Benutzer kann nach Bedarf die Karte modifizieren, wobei ihn das System unterstützen sollte.

### Ein Architekturkonzept für den Entwurf thematischer Karten

Yang, Siekierska und Taylor liefern ebenfalls einen konzeptionellen Ansatz für den automatisierten Entwurf thematischer Karten. Sie beschäftigen sich aber auch mit möglichen Softwarearchitekturen für ein entsprechendes Expertensystem sowie mit der Frage der Integration in GIS [YaST93]. In ihrer Architekturkonzeption sind Expertensystem und GIS klar voneinander getrennt und arbeiten deswegen asynchron. Sie halten diese strikte Trennung zwar für nachteilig, da so die Funktionalität und der Informationsspeicher beider Systeme sich nicht gut koppeln läßt und Informationen doppelt gehalten und verarbeitet werden müssen. Sie finden die Trennung aber notwendig, weil mit heute verfügbaren GIS keine engere Integration möglich ist.

In ihrer Systemarchitektur identifizieren sie vier Module des Expertensystems, die neu entwickelt werden müssen:

- *Informationsabfrage:* Der Benutzer wird nach den Datencharakteristika und nach der Zielsetzung der Karte gefragt. Da diese Befragung vom Expertensystem gesteuert ist, können die Fragen sehr zielgerichtet formuliert werden und es wird nur die Information erfragt, die für das aktuelle Entwurfsproblem benötigt wird.



- *Kartentypauswahl*: Aus einem Satz von Typen thematischer Karten werden solche ausgewählt, die den Datensatz ausdrücken. Dazu werden (1) geeignete Signaturen und visuelle Variablen bestimmt, (2) untersucht, ob alle Daten von einem Kartentyp ausgedrückt werden können oder ob mehrere Karten nötig sind und (3) in jedem Schritt die akzeptablen Alternativen festgehalten.
- *Kartenkombination*: Sind mehrere elementare Karten zum Ausdrücken des Datensatzes nötig, werden sie in diesem Modul kombiniert. Dabei werden verschiedene Alternativen getestet, bis eine geeignete Kombination gefunden ist.
- *Zuweisung visueller Parameter*: Den graphischen Elemente der thematischen Karte werden die visuellen Parameter wie Farbe, Größe, Form etc. zugewiesen.

#### „Augmented Intelligence“

Ein alternativer Ansatz für den automatisierten Kartenentwurf wurde von Weibel vorgeschlagen [Weib91]. Er hält die Entwicklung von vollständigen kartographischen Expertensystemen, die Generalisierung, Symbolisierung und Produktion vielfältiger Kartentypen automatisieren, für extrem aufwendig und deswegen mittelfristig für nicht realisierbar. Sein Amplified Intelligence-Ansatz setzt niedriger an: er überläßt die wichtigsten Entwurfsentscheidungen dem Benutzer und unterstützt ihn dabei mit einer Reihe von wissensbasierten Werkzeugen, die Teilaufgaben des Entwurfs ausführen.

Weibel entwirft verschiedene beispielhafte Szenarien für die Einsatzmöglichkeiten solcher Werkzeuge:

- Im Symbolisierungsmodus werden dem Benutzer verschiedene, vom Expertensystem entwickelte Alternativen in einem Browser vorgeschlagen. Dem Benutzer bleibt die Entscheidung, welche Symbolisierung letztendlich verwendet wird.
- Bei der Generalisierung sind einige enge Teilgebiete schon heute sehr gut automatisierbar, z.B. die automatische Glättung von Linien. Entsprechende Module sollen diese Teilaufgaben automatisch lösen. Dem Benutzer wird anschließend die bereits teilweise generalisierte Karte präsentiert und er erledigt die noch ausstehenden Schritte interaktiv.
- In ähnlicher Weise kann auch bei der Kartenbeschriftung verfahren werden: nur bei schwierigen Beschriftungskonflikten muß der Benutzer selbst entscheiden, die Routinearbeit wird von Softwarefunktionen realisiert.

#### 3.4.4 Wissensbasierte Symbolisierung: zwei Prototypsysteme

Nachdem verschiedene Konzepte für kartographische Entwurfssysteme diskutiert wurden, stellt dieser Abschnitt nun zwei Prototypsysteme vor. Diese Systeme lösen jeweils eine Teilaufgabe der kartographischen Symbolisierung: die Auswahl von Signaturtypen und die graphische Ausgestaltung der Signaturen.

**Wissensbasierter Entwurf kartographischer Signaturen**

Wang untersucht die automatisierte graphische Ausgestaltung der Signaturen einer thematischen Karte [Wang92]. Diese Ausgestaltung erfolgt im kartographischen Entwurfsprozeß immer erst nach der Entscheidung über die Signaturtypen. Beispielsweise muß nach der Entscheidung für die Repräsentation eines Datensatzes durch die Choroplethentechnik bestimmt werden, welche Farben den verschiedenen Datenklassen zugeordnet sein sollen. Wang hat ein entsprechendes Prolog-Programm entwickelt, das einige der häufigsten Ausgestaltungsprobleme löst:

- Form
- Form plus Farbe
- Größe punktförmiger Signaturen
- Orientierung plus Farbe
- Farbintervall von „warm“ bis „kalt“

Wang ist es gelungen, in seinem Prolog-Programm einige wichtige graphische Regeln zu implementieren. Insbesondere berücksichtigt er Konventionen bei der Farbwahl (z.B. Blau suggeriert kalt, Rot suggeriert warm) und repräsentiert auch in der Benutzerschnittstelle Farben in einem wahrnehmungsorientierten Farbmodell. Er berücksichtigt weiterhin bekannte optische Täuschungen und Unregelmäßigkeiten der visuellen Wahrnehmung, indem er beispielsweise größeren Flächen weniger intensive Farben zuweist. Der Schwerpunkt seiner Arbeit liegt insgesamt auf der farblichen Ausgestaltung von Choroplethenkarten.

**Objektorientierte, wissensbasierte Symbolisierung**

Zhan und Buttenfield [ZhBu95] beschäftigen sich mit der automatischen Symbolisierung, wobei sie sich auf thematische Karten für univariate Daten (mit einer einzigen abhängigen Variablen) beschränken. Sie argumentieren, daß das Wissen zur Symbolwahl für univariate Karten in einer objektorientierten Form effektiv repräsentiert und verarbeitet werden kann. Sie entwerfen insbesondere eine Klassenhierarchie zur Beschreibung der Symboltypen und Regeln und nutzen objektorientierte Konzepte wie die Vererbung zur Vereinfachung der Regelbasis aus. Ein Prototyp ihres Systems wurde in der Expertensystemshell Clips implementiert.

Das System entwirft Karten immer nach einem festen Schema, das aus fünf Schritten besteht:

1. Dimension des Raumbezugs (d.h. der unabhängigen Variable) erfragen: punktförmig, linienförmig oder flächenförmig?
2. Ordnungstyp der abhängigen Variablen erfragen: nominal, ordinal oder quantitativ?
3. Symboldimension bestimmen: punktförmig, linienförmig oder flächenförmig?

4. Kartentyp bestimmen
5. visuelle Variablen festlegen

In den ersten beiden Schritten wird jeweils der Benutzer befragt. In den späteren drei Schritten arbeitet das System zum Teil selbständig, d.h. es bestimmt aufgrund seiner Regeln und der gegebenen Information die benötigten Attribute. Reicht das Wissen in einem dieser Schritte nicht aus, wird der Benutzer nach weiterer Information gefragt oder ihm die Alternativen zur Entscheidung vorgelegt.

Eine offensichtliche Schwäche dieses Ansatzes ist die fehlende Effektivitätsauswertung. Es wird nur sehr rudimentär auf die verschiedenen Einflußgrößen des Kartenentwurfs eingegangen. Insbesondere besteht das Datenmodell aus nur zwei Parametern (die in den Schritten 1 und 2 erfragt werden), die Zielsetzung der Karte und weitere Bedingungen werden nicht berücksichtigt. Durch den Wegfall dieser wichtigen Parameter kann das Entwurfsproblem durch einen einfachen Entscheidungsbaum repräsentiert werden und für jeden Datenfall liegt genau eine Lösung fest. Da es keine mehrfachen Lösungen geben kann, entfällt auch die Bewertung der Effektivität.

### 3.5 Automatisierter GIS-Visualisierungsentwurf: Offene Probleme

Wie bereits in Abschnitt 2.7 dargestellt wurde, ist das Ziel dieser Arbeit die Entwicklung eines wissensbasierten Systems zur Benutzerunterstützung bei der Visualisierung in GIS. Wichtige Charakteristika eines solchen Systems sind der automatisierte GIS-Visualisierungsentwurf, die Präsentation von Visualisierungsalternativen und Erklärungen, die Möglichkeit der Modifikation von Visualisierungen und eine integrierte Möglichkeit der Weiterbildung.

Die in Abschnitt 3.2 vorgestellten Systeme für den Visualisierungsentwurf sind wichtige Ansatzpunkte, berücksichtigen aber weder die Visualisierungszielsetzung noch die spezifischen Charakteristika von Geodaten, wie die Maßstabsabhängigkeit. VISTA läßt beispielsweise nur Visualisierungen von Daten zu, die über einem regulären Gitter definiert sind. Die kartographischen Entwurfssysteme aus Abschnitt 3.4.4 berücksichtigen zwar einige Geodaten-Charakteristika, lösen aber jeweils nur einen Teilaspekt des Entwurfs. Elementare Methoden der Geodatenvisualisierung, wie die kartographische Überlagerung, werden nicht berücksichtigt. Keines der Systeme ermöglicht den Entwurf mehrerer alternativer Visualisierungen, die Abschätzung und Erklärung der Effektivität von Visualisierungen oder die systemseitige Erklärung der Entwurfsentscheidungen.

Trotz der Forderung nach einem System, das Benutzer beim Entwurf von GIS-Visualisierungen unterstützt und einiger theoretischer Arbeiten auf diesem Gebiet sind also bisher praktisch keine ernstzunehmenden Prototypsysteme für GIS entwickelt worden. Auf dem verwandten Gebiet des Visualisierungsentwurfs für nicht raumbezogene Daten waren jedoch APT, VISTA und andere Ansätze relativ erfolgreich. Die Vermutung liegt nahe, daß

die Prinzipien dieser Systeme auch für den Visualisierungsentwurf von GIS-Daten anwendbar sind.

Um die Zielsetzung dieser Arbeit, ein Benutzerunterstützungssystem für die GIS-Visualisierung zu erreichen, sollen deshalb die Entwurfsmethoden von APT und VISTA stark ausgebaut und um spezifische Funktionalität für den GIS-Visualisierungsentwurf erweitert werden. Zu den Erweiterungen muß insbesondere die Berücksichtigung von Geodatencharakteristika, Visualisierungszielsetzung und -kontext zählen. Die Entwurfs- und Kompositionsmethoden müssen den Eigenheiten der Geodaten-Visualisierung angepaßt werden und wichtige Konzepte wie Maßstab und kartographische Überlagerung umsetzen. Auch diejenigen Elemente der Umgebung, die nicht direkt zum automatisierten Entwurf zählen, also Erklärung, Modifikation und Training, müssen bei der Entwicklung der Entwurfskomponente berücksichtigt werden. Nur so ist ein wirksames Zusammenspiel aller Elemente der Umgebung gewährleistet.

# Kapitel 4

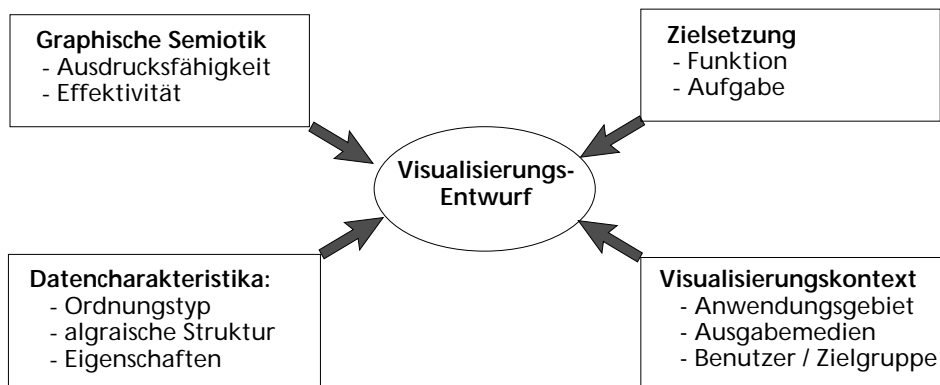
## Charakterisierung von Sachdaten und GIS-Visualisierungskontext

In Kapitel 4 wird untersucht, welche Information über Daten und GIS-Umgebung vorhanden sein müssen, damit effektive GIS-Visualisierungen automatisch entworfen werden können. Neben den Datencharakteristika haben in diesem Zusammenhang die Visualisierungszielsetzung, Benutzer und Ausgabemedien eine Bedeutung. Charakterisierungsschemas für jede dieser Informationsquellen werden aufgestellt und in einem Charakterisierungsmodell für effektive GIS-Visualisierung integriert. Dieses Modell bildet den Ausgangspunkt für die Entwicklung einer Methodik des automatisierten GIS-Visualisierungsentwurfs.

Zur Motivation der nächsten Abschnitte gibt Abb. 4-1 einen Überblick über alle wesentlichen Einflußfaktoren des Visualisierungsentwurfs. Unter diesen Faktoren sind die Gesetze der graphischen Semiotik konstant, d.h. sie sind für alle Anwendungen gleich und werden im Regelwerk der Entwurfskomponente festgehalten. Visualisierungszielsetzung, Visualisierungskontext und Datencharakteristika sind jedoch variabel und bilden die veränderlichen Informationsquellen des Visualisierungsentwurfs. Aus der Fülle an Information, die über Datensatz, Zielsetzung und Kontext existieren kann, muß der Teil ausgewählt werden, der die Ausdrucksfähigkeit und die Effektivität von Visualisierungen beeinflusst. In den folgenden Unterkapiteln werden entsprechend Charakterisierungsschemas für die wichtigsten Eingabeparameter Datensatz, Visualisierungszielsetzung und Visualisierungskontext aufgestellt.

### 4.1 Charakteristika von GIS-Sachdaten

Beim Entwerfen eines Charakterisierungsschemas für GIS-Sachdaten kann auf Schemas zurückgegriffen werden, die für verwandte Gebiete bereits existieren. APT und alle darauf aufbauenden Systeme haben den Ordnungstyp der Variablen als Hauptcharakteristikum betrachtet. Ausgangspunkt des hier vorgestellten Charakterisierungsschema für GIS-Sachdaten ist die detaillierte Charakterisierung relationaler Datensätze von Roth und Mattis [RoMa90] und die Datencharakterisierung von Senay und Ignatius [SeIg90a]. Die-



**Abb. 4-1:** Einflussfaktoren des Visualisierungsentwurfs

se Schemas wurden für die GIS-Visualisierung angepaßt und um wichtige Komponenten, insbesondere eine Charakterisierung des Raumbezugs erweitert.

### 4.1.1 Geographisches Bezugssystem

#### Maßstab

Bei raumbezogenen Daten muß der Maßstabsbereich, für den sie gültig sind, berücksichtigt werden. Da die heute verfügbaren GIS keinen kartographischen Zoom ermöglichen, darf ein Datensatz normalerweise nicht außerhalb seines Erfassungs-Maßstabsbereiches dargestellt werden. Beim Visualisierungsentwurf läßt sich der Maßstabsbereich vor allem für zwei Konsistenzprüfungen einsetzen: Zum einen sollen zwei raumbezogene Datensätze nicht überlagert werden, wenn sich ihr zulässiger Maßstabsbereich nicht überlappt. Zum anderen ist der Zoom einer Visualisierung nur im erlaubten Maßstabsbereich zulässig. Dieser Bereich wird ermittelt als die Schnittmenge der Maßstabsbereiche aller in der Visualisierung enthaltenen Datensätze.

#### Projektion

Häufig beziehen sich GIS-Daten nicht direkt auf geographische Koordinaten, d.h. Längen- und Breitengrade, sondern es liegt ein kartesisches Koordinatensystem vor. In diesem Fall wurden die geographischen Koordinaten durch eine Projektion auf die Ebene abgebildet und die Charakteristika des zugrundeliegenden Projektionsverfahrens müssen für den Visualisierungsentwurf bekannt sein. Besonders bei sehr kleinmaßstäblichen Geodaten kann bei Verwendung ungeeigneter, nicht der Projektion entsprechender Visualisierungstechniken ein stark verzerrtes Bild des Datensatzes entstehen.

Wichtige Eigenschaften der Projektion sind Formtreue, Flächentreue, Längentreue und Winkeltreue (Konformität) [Bill96]. Die für Darstellungen des ganzen Globus sehr häufig verwendete Mercator-Projektion ist beispielsweise formtreu, d.h. sie gibt die Form der Kontinente im wesentlichen korrekt wieder.

Flächentreue der Projektion ist wichtig, wenn die Dichte eines Phänomens bewertet werden soll. Ist z.B. eine Punktkarte nicht flächentreu, dann ist der Dichteeindruck, insbesondere in den Randbereichen, fehlerhaft. Es kann nicht mehr entschieden werden, ob Punkthäufungen auf Häufungen des Phänomens oder auf Verzerrungen der Projektion zurückzuführen sind. Andererseits ist bei allen GIS-Visualisierungen, die für die Lageplanung verwendet werden, die Längentreue eine wichtige Voraussetzung. Nur unter dieser Annahme lassen sich Abstände, Erreichbarkeit und Wechselwirkungen eines Standorts mit seiner Umgebung richtig bewerten.

Generell sollten alle Visualisierungstechniken, die Werteinterpolationen einschließen, diese Interpolation im sphärischen oder geoiden Koordinatensystem und nicht in den projizierten, kartesischen Koordinaten durchführen (siehe hierzu auch [MacE94a], Seite 68). Solche Interpolationen fallen besonders bei kontinuierlichen Daten, z.B. im Zusammenhang mit der Isarithmentchnik an. Im Falle kleinmaßstäblicher, kontinuierlicher Daten sollten also die ursprünglichen geographischen Koordinaten zur Verfügung stehen, oder die Angabe der Projektionsmethode und ihrer Parameter eine Rückprojektion ermöglichen.

#### 4.1.2 Relationen und Variablen

Zur Diskussion der Struktur von Datensätzen gehen wir der Einfachheit halber vom relationalen Datenmodell aus. Die große Mehrzahl praktischer GIS basiert nach wie vor auf diesem Datenmodell. Selbst in Systemen mit objektorientierter Datenhaltung sind quantitative Sachdaten häufig in Form von Relationen abgelegt, denn ihre Struktur wird so am besten wiedergegeben. Zunächst sollen hier einige Begriffe informell eingeführt werden. Eine exakte Definition dieser Begriffe folgt dann zu Beginn von Kapitel 5.

Eine Relation besteht aus einer Anzahl von Variablen (in der Datenbankterminologie auch *Attribute* genannt). Sind einige der Variablen durch andere bestimmt, werden die ersteren als abhängig und die letzteren als unabhängig bezeichnet. Treten keine Abhängigkeiten zwischen Variablen auf, werden sie alle unabhängig genannt. Die Daten sind in der Relation in Tupel strukturiert, wobei jedem Element des Tupels (und damit auch jeder Stelle im Tupel) eine Variable der Relation entspricht.

Alle Relationen einer Anwendung werden konzeptionell zu einer Datenbank zusammengefaßt. Ein Datensatz besteht aus einer oder mehreren Relationen der Datenbank. Der Datensatz kann statt den vollständigen auch Teilrelationen enthalten, wenn beispielsweise Variablen weggelassen sind, oder aus allen Tupeln der Relation einige ausgewählt wurden.

##### **Strukturelle Eigenschaften der Relation**

Für jede Relation ist zunächst die Unterscheidung ihrer Variablen in unabhängige und abhängige Variablen wichtig. Weitere für den Visualisierungsentwurf interessante Eigenschaften von Relationen sind Kardinalität, Vollständigkeit und Eineindeutigkeit:

- *Kardinalität:* Die Anzahl von Wertetupeln der abhängigen Variablen, auf die ein Tupel der unabhängigen Variablen jeweils abgebildet wird. Eine Kardinalität von 1 impliziert, daß es sich bei der Relation um eine Funktion handelt, eine Vorbedingung für zahlreiche Techniken der thematischen Kartographie. Es können auch variable Kardinalitäten auftreten.
- *Vollständigkeit:* Ist die Relation *vollständig*, dann wird jedes Tupel der unabhängigen Variablen auf mindestens ein Wertetupel abgebildet.
- *Eindeindeutigkeit:* Diese Eigenschaft bedeutet, daß jedes Tupel der unabhängigen Variablen auf genau ein Wertetupel abgebildet wird.

### Algebraische Eigenschaften der Relation

Häufig sind Relationen so aufgebaut, daß auch zwischen abhängigen Variablen funktionale Zusammenhänge bestehen. So könnte z.B. eine abhängige Variable *Gesamtbevölkerung* die Summe zweier weiterer abhängiger Variablen *Männer* und *Frauen* sein. Oder aber die Summe von *A* und *B* ist nicht explizit als weitere abhängige Variable in der Relation vorhanden, *A* und *B* sind aber Teile eines Ganzen. Diese Information sollte, genau wie die über funktionale Abhängigkeiten zwischen Variablen, in den Datencharakteristika enthalten sein. Dann können Visualisierungsmethoden, wie z.B. Tortendiagramme, gewählt werden, die diese Zusammenhänge besonders gut wiedergeben.

Mehrere Variablen können auch Teile eines komplexeren Datentyps sein und deswegen spezielle Visualisierungstechniken erfordern. Ein einfaches Beispiel hierfür sind die abhängigen Variablen *Temperaturminimum* und *Temperaturmaximum*. Werden sie mit einem Balkendiagramm visualisiert, sollten die beiden Variablen immer zusammen auf einem Balken dargestellt werden.

### Ordnungstyp der Variablen

Für den Visualisierungsentwurf ist das Hauptmerkmal einer Variable sicherlich ihr Ordnungstyp. Nominale Datensätze, z.B. Landnutzungsdaten, erfordern andere Visualisierungstechniken als quantitative Daten, bei denen ein kontinuierliches Werteintervall abgebildet wird. Folgende drei Ordnungstypen werden beim Visualisierungsentwurf unterschieden:

- *nominal:* Der Wertebereich der Variablen hat keine natürliche Ordnung. Ein Beispiel für nominale Variablen ist die Landnutzungsart.
- *ordinal:* Der Wertebereich der Variablen ist linear geordnet. Ein Beispiel sind die Waldschadensklassen I, II, III und IV.
- *quantitativ:* Der Wertebereich der Variablen ist linear geordnet und hat ein präzises, numerisches Maß. Ein Beispiel ist die Niederschlagsmenge in Liter/Quadratmeter.

Im GIS-Bereich ist es sinnvoll, einen weiteren „Ordnungstyp“ einzuführen:



- *geographisch*: Der Wertebereich der Variablen ist durch die geographische Lage geordnet. Der geographische Bezug kann der 2D- oder 3D-Raum oder auch ein spezielles (z.B. geoides) Koordinatensystem sein. Beispiele für solche Variablen sind Koordinaten, Vektorobjekte und geographische Referenzen in Form von Adressen.

Es handelt sich hier zwar nicht um einen neuen Ordnungstyp im Sinne der Ordnungstheorie (dort ist er zum quantitativen Ordnungstyp äquivalent), aus praktischen Gründen sind geographische Variablen aber eine sehr nützliche Erweiterung der klassischen Dreiteilung der Ordnungstypen. Der Raumbezug spielt im GIS-Kontext eine herausragende Rolle: geographische Variablen werden fast immer durch die visuelle Variable *2D-Raum* bzw. *3D-Raum* dargestellt.

Mit einer ähnlichen Argumentation ließe sich auch ein temporaler „Ordnungstyp“ fordern, temporale Daten lassen sich aber gut als Spezialfall ordinaler oder quantitativer Daten modellieren (siehe hierzu auch Abschnitt 4.1.4).

### 4.1.3 Geographische Variablen

#### Dimension und Bezugssystem

Geographische Variablen können in verschiedenen geographischen Bezugssystemen definiert sein. Das Bezugssystem läßt sich an der Dimension und Bedeutung der Koordinaten festmachen. Die vier häufigsten Koordinatentypen in GIS sind:

- *geographisch*: Die Koordinaten geben Längen- und Breitengrade an.
- *2D kartesisch*: Die geographischen Koordinaten wurden projiziert auf ein kartesisches System mit  $x$ - und  $y$ -Werten (bei der Gauß-Krüger-Projektion Rechts- und Hochwerte genannt).
- *2,5D kartesisch*: Zu den projizierten  $x$ - und  $y$ -Werten des 2D-Systems kommt noch ein  $z$ -Wert, der die Höhe des Punktes über Normal-Null angibt.
- *3D kartesisch*: Die Koordinatenwerte sind nicht projiziert. Statt dessen wird ein Punkt (nicht notwendigerweise auf der Erdoberfläche) als Nullpunkt definiert und dort ein dreidimensionales, kartesisches Koordinatensystem verankert.

In manchen Fällen wird auch mit Sonderkoordinatensystemen gearbeitet, wie z.B. mit Polar- oder Kugelkoordinaten.

Sollen Geodaten mit unterschiedlichen Bezugssystemen zusammen verarbeitet oder visualisiert werden, dann ist neben dem Koordinatentyp auch die Angabe der Referenzpunkte und der zugrundeliegenden Projektion erforderlich.

#### Rasterdaten

Unter Rasterdaten versteht man im GIS-Kontext Daten, die für jede Zelle in einem regelmäßigen Gitter vorliegen, wie z.B. Satellitenbilder. Mittels Rasterdaten werden vor allem

kontinuierliche Phänomene erfaßt. Für die Verarbeitung und Visualisierung von Rasterdaten muß die Form des zugrundeliegenden Gitters bekannt sein:

- *reguläre Gitter:* Alle Gitterzellen sind rechteckig und gleich groß. Es genügt die Angabe des Gitterabstandes in  $x$ -, in  $y$ - und in  $z$ -Richtung.
- *Gitter mit Sonderform:* Hier muß zusätzlich die Form des Gitters angegeben werden. Gitter mit Sonderformen sind bei einigen numerischen Simulationen erforderlich, z.B. bei der Finit-Element-Analyse.
- *Unregelmäßige Koordinaten:* Sie treten beispielsweise bei Lotungen in der Schifffahrt oder bei Messungen mit Meßwagen und GPS-Positionsbestimmung auf. Sind die Meßorte unregelmäßig, muß zu jedem Meßwert auch eine eigene Position festgehalten werden. Die Daten werden dann manchmal auch als Vektordaten betrachtet.

### Vektordaten (Geo-Objektdaten)

Vektordaten (besser als Geo-Objektdaten bezeichnet) beziehen sich im Gegensatz zu Rasterdaten nicht auf ein festes Gitter. Statt dessen gehört zu jedem Objekt auch eine Geometrie (bei punkthaften Objekten eine Position). Hauptunterscheidungsmerkmal für geographische Variablen vom Typ Geo-Objekte ist die räumliche Dimension der Objekte. Sie können punktförmig, linienförmig, flächenförmig oder volumenförmig sein. Eine Variable kann allerdings auch Geo-Objekte unterschiedlicher Dimension umfassen.

### Geo-Referenzdaten

Geo-Referenzen sind Verweise auf Geodaten. So verweist z.B. eine Adresse mit den Komponenten Straße und Hausnummer auf ein Geo-Objekt (Gebäude) oder mehrere Geo-Objekte, die punkt-, flächen- oder volumenförmig modelliert sein können. Zu Geo-Referenzvariablen sollte also ebenfalls die Dimension bekannt sein.

## 4.1.4 Attributvariablen

Nominale, ordinale und quantitative Variablen werden unter dem Begriff Attributvariablen zusammengefaßt.

### Datentyp

Neben dem Ordnungstyp lassen sich Attributvariablen auch anhand ihres Datentyps unterscheiden. Variablen von jedem der drei Ordnungstypen können in verschiedenen Datentypen auftreten, die in der Regel auch spezifische Visualisierungstechniken nahelegen.

Werte nominaler Variablen treten in folgenden drei Datentypen auf:

- *Aufzählungstyp:* Es gibt eine feste, beschränkte Menge (Aufzählung) von Werten. Zur Speicherung und Verarbeitung in GIS sind die Werte häufig durch Schlüssel repräsentiert, z.B. Roggen  $\cong$  1, Weizen  $\cong$  2, Hafer  $\cong$  3, etc.

- *Zeichenkette:* Die Wertemenge ist nicht beschränkt und eine Schlüsselzuordnung folglich nicht möglich. Beispiel: Der Name einer Grundstückseigentümerin.
- *Media-Referenz:* Solche Referenzen oder Links werden ebenfalls als Zeichenkette repräsentiert, haben jedoch eine spezielle Semantik: Sie geben den Typ und den Zugriffspfad eines Multimedia-Objektes wieder. Media-Referenzen treten mit der Entwicklung multimedialer und World-Wide-Web-basierter GIS immer häufiger auf. Im WWW sind sie durch die URL (universal resource locator) standardisiert.

Zu den ordinalen Variablen gehören folgende zwei Datentypen:

- *geordnete Liste:* Hier gibt es wieder eine feste, beschränkte Menge von Werten, die jedoch zusätzlich eine Ordnung besitzen. Die Repräsentation erfolgt ebenfalls über Schlüssel.
- *ganze Zahl*

Quantitative Variablen schließlich treten nur in Form von Zahlen auf.

### **Zeit als Variable**

Wie schon weiter oben erwähnt wurde, kann die Zeit als Variable entweder ordinal (z.B. zeitliche Ereignisse) oder quantitativ (z.B. Meßzeitpunkte) sein. Der Ordnungstyp hängt hier also von der Bedeutung der Variablen ab.

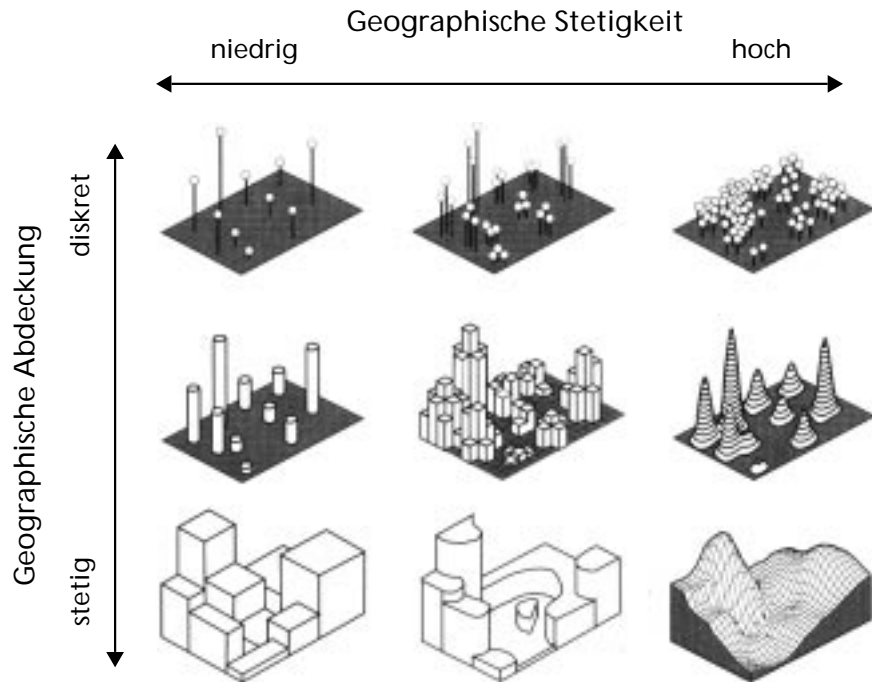
Eine Sonderstellung nimmt die Zeit vor allem in den Fällen ein, wo sie als unabhängige Variable auftritt, was in GIS-Anwendungen auch häufig der Fall ist. Dann gibt es nämlich eine natürliche Abbildung der temporalen Datenvariable in die visuelle Variable „Darstellungszeit einer Animationssequenz“. Temporale Variablen sollten demnach als solche gekennzeichnet sein.

### **Geographische Abdeckung und Stetigkeit**

Um einen Datensatz in effektiver Form visuell darzustellen, muß ausreichende Information über das durch die Daten repräsentierte natürliche Phänomen vorhanden sein. Abb. 4-2 zeigt die wesentlichen zwei Kriterien: die geographische Abdeckung und die geographische Stetigkeit des Phänomens.

Die geographische Abdeckung läßt sich zum Beispiel durch folgende drei Stufen ausdrücken:

- *diskret (isoliert):* Das Phänomen tritt nur an isolierten Punkten des Bezugsgebiets auf. Beispiel: das zur Variable „Anzahl Arbeiter in der Asbestverarbeitung“ gehörende Phänomen „asbestverarbeitender Betrieb“ tritt nur an isolierten Punkten auf.



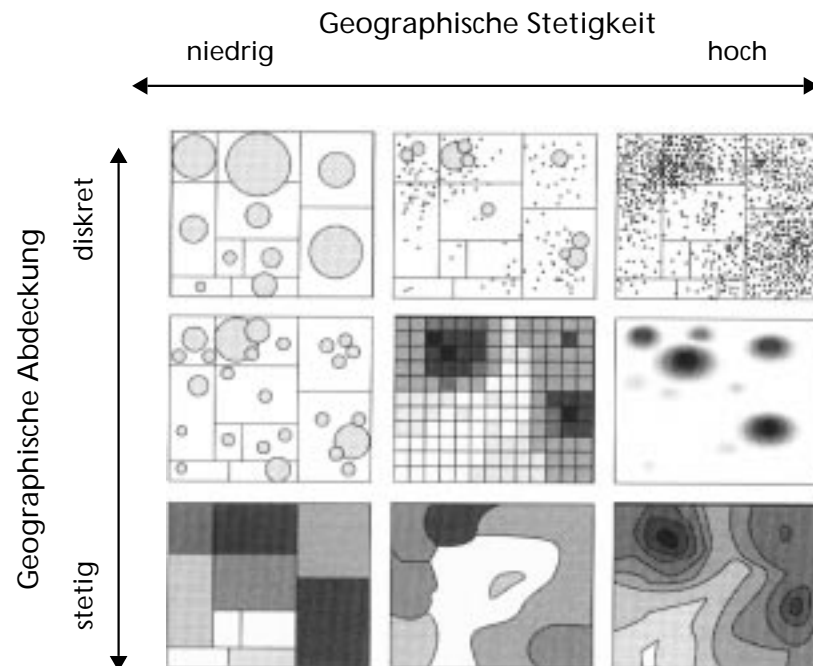
**Abb. 4-2:** Geographische Abdeckung und geographische Stetigkeit, nach [MacE94a]

- *lückenhaft:* Das Phänomen tritt flächenhaft, aber nicht an jeder Stelle des Bezugsgebiets auf. Beispiel: das zur Variable „Anzahl Cholerafälle“ gehörende Phänomen „Cholera“ tritt nur in bestimmten Regionen auf.
- *kontinuierlich:* Das Phänomen tritt an jeder Stelle des Bezugsgebiets auf. Beispiel: das zur Variable „mittlere Tagestemperatur“ gehörende Phänomen „Wetter“ tritt kontinuierlich auf.

Im Gegensatz zur geographischen Abdeckung ist die Stetigkeit eines Phänomens nur für quantitative Variablen definiert. Bei stetigen Phänomenen kann bei einer kleinen Variation der Position auch der Variablenwert nur begrenzt variieren. Die Temperatur ist eine solche stetige Variable. Im Gegensatz dazu treten bei unstetigen Phänomenen auch bei kleiner Variation der Position mitunter große Variationen des Variablenwertes auf.

Abdeckung und Stetigkeit hängen immer auch vom Maßstab ab, unter dem ein Phänomen betrachtet bzw. erfaßt wird. Was bei einem großen Maßstab aufgrund zahlreicher lokaler Variationen unstetig ist, kann im kleinen Maßstab durchaus stetig erscheinen. Betrachtet man beispielsweise die Bevölkerungsdichte auf der Kreisebene, treten an den Siedlungsgrenzen Unstetigkeiten auf, während die Dichte im kleinen Maßstab einer Europakarte relativ stetig erscheint.

Abdeckung und Stetigkeit eines Phänomens üben beide einen starken Einfluß auf den Eignungsgrad verschiedener Visualisierungstechniken aus. Dieser Zusammenhang wird in Abb. 4-3 exemplarisch dargestellt.



**Abb. 4-3:** Darstellungsformen bei unterschiedlichen Charakteristika, nach [MacE94a]

### Numerische und statistische Eigenschaften

Bei quantitativen Variablen haben einige numerische und statistische Attribute einen bedeutenden Einfluß auf den Visualisierungsentwurf. Gesichertes Wissen über numerische Eigenschaften der Variablenwerte ist Bedingung für die Verwendbarkeit einiger Visualisierungstechniken. Beispielsweise kann die Choroplethentechnik nicht angewendet werden, wenn die Daten absolute Werte sind. Die folgende Liste gibt die wichtigsten Erscheinungsformen quantitativer Werte wieder, wie sie in GIS-Datenbanken vorkommen:

- *Anzahlen:* Absolute natürliche Zahlen, die Mächtigkeiten von Mengen darstellen. Anzahlen können abhängig oder auch unabhängig von der Fläche sein, je nach der Natur des betrachteten Phänomens.
- *Maßzahlen:* Absolute Zahlen, die Werte eines Maßes (z.B. des Abstandes) repräsentieren. Auch Maßzahlen können abhängig oder unabhängig von der Fläche sein. Zu Maßzahlen sollte auch die Information über die zugehörige Maßeinheit vorhanden sein.
- *aggregierte Zahlen:* Die Werte sind Anzahlen oder Maßzahlen, die über eine Fläche summiert worden sind. Sie sind damit implizit abhängig von der Fläche.
- *Verhältniszahlen:* Die Zahlen wurden durch Division mit einer weiteren Größe normiert.
- *Dichten:* Anzahlen, oder aggregierte Anzahlen, die durch den zugehörigen Flächeninhalt dividiert wurden. Sie sind damit unabhängig von der Fläche.

- *Koordinaten:* Die Werte sind nicht als Maßzahlen, sondern als Positionen in einem Koordinatensystem aufzufassen. Dies ist eher eine semantische als eine numerische Eigenschaft, und viele Größen können je nach Bedeutung entweder Koordinaten oder Maßzahlen sein.

Statistische Eigenschaften von Variablen hängen vom Ordnungstyp ab. Für nominale Variablen ist nur die Werteanzahl definiert. Für quantitative und (mit Einschränkungen) auch für ordinale Variablen gibt es noch eine Reihe weiterer Attribute. Sie können zum Teil automatisch aus der Wertemenge der Variablen ermittelt werden:

- *Definitionsbereichminimum und -maximum*
- *Werteminimum und -maximum:* Im Gegensatz zum Definitionsbereich der Variablen werden hier die Extremwerte betrachtet, die im Datensatz auch vorkommen.
- *Verteilungsart:* Ist die Verteilungsart einer Variable bekannt, kann diese Information zur Bestimmung einer effektiven Datenklassifizierung verwendet werden.
- *Mittelwert und Varianz:* Vor allem bei Normalverteilungen. Zu anderen Verteilungsarten sind deren spezifische Parameter interessant.

#### 4.1.5 Datenqualität

Das Thema Datenqualität hat besonders in den letzten fünf Jahren in der Kartographie und in der GIS-Forschung große Beachtung erlangt. Informationen zur Qualität von Geodaten sind sehr wichtig, da der Erfassungsaufwand und die Kosten solcher Daten immens sind und sie deswegen nicht immer aktuell und in höchster Qualität vorliegen. Fehlende Datensätze, fehlerhafte Daten und zweifelhafte Werte können aber eine Darstellung verfälschen und fehlerhafte Interpretationen zur Folge haben. Eine sehr detaillierte Diskussion der Qualität von Geodaten findet sich in [MacE94a, S. 74-85], sowie in [Chri91].

Die elementarste Information zur Datenqualität ist eine Unterscheidung des Nullwertes von nicht gemessenen oder nicht vorhandenen Werten. Werden diese beiden Fälle in der Datenbank nicht unterschieden, zeigt beispielsweise eine Choroplethenkarte in beiden Fällen die Farbe der niedrigsten Kategorie an. Nicht gemessene Werte sollten jedoch gar nicht abgebildet werden, und in der Choroplethenkarte z.B. weiß bleiben.

##### **Zuverlässigkeit**

Die zwei wesentlichen Faktoren der Datenqualität sind Zuverlässigkeit und Genauigkeit. Unter dem Begriff Zuverlässigkeit wird bewertet, wie gut die Messung bzw. die Erfassung das Phänomen wiedergibt. Aspekte der Zuverlässigkeit sind deswegen die Voraussetzungen bzw. Bedingungen des Meßverfahrens sowie die auftretenden Meßfehler. Im einzelnen sollten Informationen zu folgenden Aspekten vorliegen:

- *Vollständigkeit:* Ist das Phänomen in den vorliegenden Daten vollständig erfaßt oder nur repräsentativ bzw. stichprobenartig?

- *Aktualität:* Wie aktuell sind die Daten? Oder: Wie groß ist die Distanz zwischen Erfassungszeitraum und dem Zeitraum, den die Daten repräsentieren sollen?
- *Attributfehler:* Mit welchen Meß- bzw. Erfassungsfehlern sind die Attributdaten behaftet? Fehler können explizit bekannt sein, oder durch Toleranzen bzw. Wahrscheinlichkeiten beschrieben werden.
- *räumliche Fehler:* Neben der Repräsentation eines Phänomens durch einen Attributwert kann auch sein räumlicher Bezug fehlerhaft erfaßt worden sein. Für die Repräsentation des räumlichen Fehlers gilt das gleiche wie für Attributfehler.

### Genauigkeit

Unter dem Oberbegriff Genauigkeit fragt man nicht nach Fehlern bei der Erfassung sondern nach der Repräsentation der erfaßten Daten im GIS. Aspekte der Genauigkeit sind die numerische Präzision der Daten und der Detaillierungsgrad:

- *numerische Präzision:* Auf wieviele Stellen genau wurden die Daten gemessen bzw. erfaßt?
- *Auflösung:* In welcher Detaillierung wurden die Daten erfaßt? Hier kann als Antwort die räumliche und/oder die zeitliche Grundeinheit stehen.
- *Aggregation:* Wurden die Daten aggregiert und dadurch die räumliche (zeitliche) Auflösung verringert? Als Antwort kann die Aggregationsbasis stehen.

### Charakterisierungsbasis

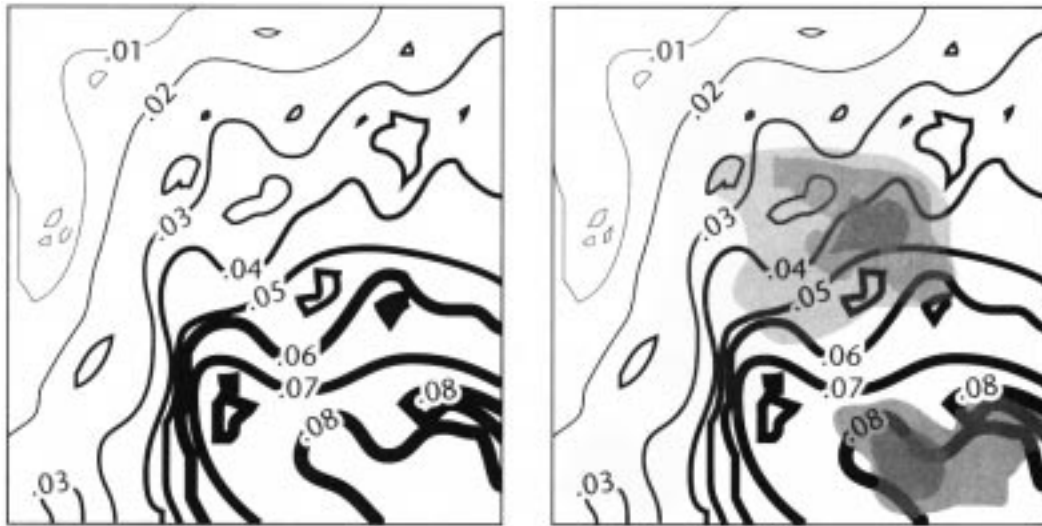
Wie bei den Geodaten selbst, sind auch bei der Qualitätsinformation verschiedene Detaillierungsstufen vorstellbar. Aspekte der Qualitätsinformation können beispielsweise für jedes Wertetupel einer Relation aber auch für die Relation als Ganzes vorliegen. Es gibt jedoch eine relativ klare Zuordnung von Qualitätsaspekten zu den Detaillierungsstufen. Beispielsweise wirken sich Meßtoleranzen und Fehlerwahrscheinlichkeiten auf alle Werte einer Variablen gleichermaßen aus, während Erfassungslücken in jedem Tupel möglich sind. Die Datenqualität sollte also auf folgender Basis charakterisiert werden:

- *pro Wert:* Unterscheidung von Nullwerten und Erfassungslücken, Zuverlässigkeit: Klassen oder Prozentzahlen.
- *pro Variable:* Meßtoleranzen, Fehlerwahrscheinlichkeit, Aggregationsbasis.
- *pro Relation:* Erfassungsdatum, Erfassungsart (Messung, Stichprobe, repräsentativ oder vollständig).
- *pro Datenbank:* Aktualität und Zuverlässigkeit der geographischen Basisdaten.

### Visualisierungsformen

Wie läßt sich die Qualität eines Datensatzes nun in seiner Visualisierung darstellen? MacEachren diskutiert auch hierfür verschiedene Möglichkeiten. Treten in der Qualität eines

Datensatzes große Schwankungen auf, kann man die Qualität einer Datenvariablen als eigenständige Variable auffassen und in einer Überlagerung zusammen mit der Datenvariablen visualisieren. Abb. 4-4 zeigt ein Musterbeispiel für diese Technik, bei dem die Unsicherheit durch verschiedene Graustufen (dunkler bedeutet unsicherer) überlagert ist.



**Abb. 4-4:** Visualisierung von Unsicherheit durch Überlagerung eines „Grauschleiers“ [MacE94a]

Diese Technik läßt sich nicht immer anwenden, insbesondere dann nicht, wenn mehrere Variablen mit voneinander unabhängiger Qualität zusammen visualisiert werden sollen. In diesem Fall lassen sich auch visuelle Variablen direkt für die Repräsentation von Unsicherheit einsetzen. Zwei der visuellen Variablen werden von Betrachtern mit Unsicherheit in Verbindung gebracht. Dazu zählt einerseits die Farbsättigung, da intensive Farben, z.B. ein sattes Rot autoritativer wirken als ungesättigte Farben wie Rosa. MacEachren führt alternativ auch den Fokus oder die Schärfe eines kartographischen Symbols als neue visuelle Variable ein, die sich zur Repräsentation von Unsicherheit eignet (siehe Abb. 4-5). Man kann argumentieren, daß die Schärfe an dieser Stelle sogar besser geeignet ist, da die Identifikation von Unschärfe mit Unsicherheit bei Betrachtern meist intuitiv gegeben ist. Sättigung kann nämlich auch anders interpretiert werden, z.B. als Intensität eines Phänomens. Schärfe, Sättigung und andere visuelle Variablen werden vereinzelt schon in



**Abb. 4-5:** Die visuelle Variable „Fokus“ bei punktförmigen, linienförmigen und flächenförmigen Objekten [MacE94a]



prototypischen GIS zur Visualisierung von Unsicherheit eingesetzt [JiBO96, FEDM96, WeGG96].

#### 4.1.6 Ein formales Charakterisierungsschema

Für die Speicherung und Verarbeitung in einem Informationssystem müssen die bisher eher unformal diskutierten Datencharakteristika in einem formalen, eindeutigen Schema zusammengefaßt werden. Ein solches Schema kann nicht den Anspruch auf Vollständigkeit und Allgemeingültigkeit erheben, da spezielle Anwendungen ihre Ergänzungen zu diesem Schema machen müssen, wenn sie andere Aspekte berücksichtigen sollen. Das vorgestellte Schema dient aber als Ausgangspunkt für die meisten praktischen GIS-Anwendungen.

Eine Schwierigkeit beim Aufstellen eines konkreten Schemas ist, einerseits nicht überzuspezifizieren, damit Flexibilität und Konfigurierbarkeit erhalten bleiben. Andererseits muß es so detailliert, aussagekräftig und konkret sein, daß die Methodik für den Visualisierungsentwurf darauf aufsetzen kann.

Das Charakterisierungsschema hat eine LISP-ähnliche Syntax und wird unter Verwendung der BNF-Notation definiert. Für eine LISP-ähnliche Syntax sprechen in diesem Zusammenhang mehrere Punkte: sie ist sehr kompakt, leicht verständlich und kommt praktisch ohne Schlüsselworte aus. Daher verwenden auch fast alle regelbasierten Expertensysteme eine ähnliche Syntax für die Repräsentation von deklarativem und strukturiertem Wissen. Die BNF-Notation hat ähnliche Vorteile und ist in der Informatik und den verwandten Disziplinen sehr verbreitet. Leider werden die Elemente und Symbole dieser Notation nicht ganz einheitlich verwendet. Deshalb ist hier zunächst eine kurze Definition der benötigten BNF-Elemente notwendig.

##### BNF-Notation

Eine BNF-Syntax besteht aus einer oder mehreren Ersetzungsregeln der Form

$$\textit{Nonterminalsymbol} = \textit{Ausdruck}$$

Elemente eines Ausdrucks sind:

- *Nonterminalsymbole*: Dies sind Verweise auf weitere Ersetzungsregeln. Ein Satz der Syntax darf keine Nonterminalsymbole enthalten, sie müssen alle ersetzt werden. Nonterminalsymbole werden kursiv dargestellt.
- *Terminalsymbole*: Diese Symbole dürfen in einem Satz der Syntax enthalten sein. Terminalsymbole werden sans-serif dargestellt, z.B.: Variable.
- [ *Ausdruck* ]: Der Ausdruck ist optional, d.h. er darf in den Satz eingesetzt, kann aber auch ausgelassen werden.

- { *Ausdruck\_1* | *Ausdruck\_2* | ... | *Ausdruck\_n* }: Die verschiedenen Ausdrücke sind Alternativen, d.h. genau einer der Ausdrücke darf in den Satz eingesetzt werden.
- *Ausdruck*\*: Der Ausdruck kann beliebig oft eingesetzt werden, kann aber auch ganz ausgelassen werden.
- *Ausdruck*<sup>+</sup>: Der Ausdruck kann beliebig oft eingesetzt werden, muß aber mindestens einmal eingesetzt werden.
- *Ausdruck\_1* *Ausdruck\_2*: Die beiden Ausdrücke bilden eine Sequenz und müssen hintereinander stehen.
- { *Ausdruck* }: Ein Ausdruck kann geklammert sein, um den Anwendungsbereich eines Operators zu kennzeichnen.

### Datensatz

Die Charakterisierung eines Datensatzes umfaßt die Beschreibung seiner Relationen, sowie Standardangaben über das geographische Bezugssystem und die Datenqualität.

```

DataSet      = ( data-set
                  ( name DBName )
                  ( relations Relation+ )
                  GeoSystem
                  Quality
                )

```

Zu den Standardangaben über das geographische Bezugssystem gehören Koordinatensystem und -dimension, Maßstabsbereich und (optional) die Projektion. Diese Standardangaben gelten für den gesamten Datensatz, sofern sie nicht von relationsspezifischen oder variablenspezifischen Angaben überschrieben werden.

```

GeoSystem    = ( dimension { 2 | 3 } )
                  ( coordinate-system { cartesian | geoid | polar | ... } )
                  ( scale Number to Number )
                  [ ( projection [ Name ] [ equal-form ] [ equal-area ]
                      [ equal-distance ] [ equal-direction ] ) ]

```

Auch die Qualitätsangaben gelten standardmäßig für den ganzen Datensatz, soweit sie nicht von spezifischen Angaben überschrieben werden. Es handelt sich um Zeitraum und Art der Erfassung, sowie optional um Erfassungstoleranzen und -fehlerwahrscheinlichkeiten.

```

Quality      = ( acquisition-period Date Date )
                  ( acquisition { complete | representative | sample
                                | estimate } )
                  [ ( measuring-tolerance Number ) ]
                  [ ( error-probability Number ) ]

```

**Relation**

Die Charakterisierung einer Relation umfaßt unabhängige und abhängige Variablen, Eigenschaften wie Kardinalität und Vollständigkeit, algebraische Abhängigkeiten zwischen den Variablen, sowie optionale spezifische Angaben über das Geo-Bezugssystem und die Qualität.

Eine Relation kann mehrere unabhängige und mehrere abhängige Variablen enthalten, muß aber mindestens eine unabhängige Variable enthalten. Es wird zwischen geographischen, nominalen, ordinalen und quantitativen Variablen unterschieden.

Strukturelle Abhängigkeiten zwischen Variablen sind die Intervalleigenschaft (zwei Variablen sind Unter- und Obergrenze eines Intervalls) sowie die Partitioneigenschaft (eine Anzahl von Variablen bilden eine vollständige Partition eines Oberbegriffs - zum Beispiel: Arbeiter, Angestellte, Beamte und Selbständige als Partition der Berufstätigen).

Funktionale Abhängigkeiten zwischen Variablen werden als mathematischer Ausdruck angegeben, z.B. ( functional Density ( / Amount Area ) )

```

Relation      = ( relation
                  ( name RelName )
                  ( independent Variable+ )
                  [ ( dependent Variable+ ) ]
                  ( cardinality { Number | variable } )
                  [ ( properties RelProperty+ ) ]
                  { ( structure { interval | partition }
                     VarName VarName* ) }*
                  { ( functional VarName MathExpression ) }*
                  [ GeoSystem ]
                  [ Quality ]
                )

RelProperty   = { complete | not-complete | unique | not-unique }

Variable      = { GeoVariable | NomVariable | OrdVariable
                  | QuantVariable }

```

**Variablen**

Eine Geo-Variable kann vom Datentyp Koordinate, Geo-Objekt oder Geo-Referenz sein. Die Charakterisierung kann optionale spezifische Angaben über das Geo-Bezugssystem und die Qualität enthalten.

```

GeoVariable   = ( variable
                  ( name VarName )
                  ( order-type spatial )
                  ( data-type { coordinate | object | reference } )
                  [ GeoSystem ]
                )

```

```

    [ Quality ]
  )

```

Eine nominale Variable kann vom Datentyp Aufzählung (ungeordnet), Zeichenkette oder Media-Referenz sein. Im Falle einer Aufzählung muß die Charakterisierung auch den Wertebereich angeben. Wie bei ordinalen und quantitativen Variablen spezifiziert die Charakterisierung die geographische Stetigkeit und optional auch die Qualität.

```

NomVariable = ( variable
    ( name VarName )
    ( order-type nominal )
    ( data-type { enumeration | string | media-reference } )
    [ ( value-range Value+ ) ]
    ( spatial-continuity { high | medium | low } )
    [ Quality ]
  )

```

Eine ordinale Variable kann vom Datentyp Aufzählung oder Zahl sein. Aufzählungen sind im ordinalen Fall geordnet und es wird der Wertebereich angegeben. Zahlen können auch Zeitpunkte repräsentieren, in diesem Fall wird die temporale Eigenschaft in der Charakterisierung wiedergegeben.

```

OrdVariable = ( variable
    ( name VarName )
    ( order-type ordinal )
    ( data-type { enumeration | number } )
    [ ( properties is-time ) ]
    { ( value-range Value+ )
      | ( value-bound Number Number ) }
    ( spatial-continuity { high | medium | low } )
    [ Quality ]
  )

```

Quantitative Variablen haben immer den Datentyp Zahl, können je nach ihrer Semantik aber Zeitpunkte, Anzahlen, Verhältniszahlen, Dichten oder Aggregationszahlen sein. Es werden die Minima und Maxima des Werte- und des Definitionsbereichs angegeben. Außerdem die Präzision der numerischen Darstellung (Vor- und Nachkommastellen), Mittelwert, und optional die Standardabweichung und die Verteilungsart.

```

QuantVariable = ( variable
    ( name VarName )
    ( order-type quantitative )
    ( data-type number )
    [ ( properties [ is-time ] [ is-amount ] [ is-ratio ]
      [ is-density ] [ is-aggregate ] ) ]
    ( value-bound Number Number )
    [ ( value-domain-bound Number Number ) ]
  )

```

```

        ( precision Number Number )
        ( distribution Number [ Number ] [ Name ] )
        ( spatial-continuity { high | medium | low } )
        ( spatial-variation { smooth | medium | abrupt } )
        [ Quality ]
    )

```

#### Beispiel 4-1 Ein Klimadatensatz

Wir betrachten als Beispiel die Charakterisierung eines Klimadatensatzes. Er besteht aus einer einzigen raumbezogenen Relation *Klima* mit den unabhängigen Variablen *Land* und *Monat*, sowie den abhängigen Variablen *Niederschlag* und *Temperatur*. Land ist räumlich, Monat ordinal und die beiden abhängigen Variablen sind jeweils quantitativ.

```

(data-set
  (name Klima1997)
  (relations
    (relation
      (name Klima)
      (independent
        (variable
          (name Bezirk)
          (order-type spatial)
          (data-type reference)
        )
        (variable
          (name Monat)
          (order-type ordinal)
          (data-type enumeration)
          (properties is-time)
          (value-range Januar Februar März ... Dezember)
        )
      )
    )
    (dependent
      (variable
        (name Niederschlag)
        (order-type quantitative)
        (data-type number)
        (properties is-density)
        (value-bound 0 113)
        (precision 3 0)
        (distribution 73)
        (spatial-continuity high)
        (spatial-variation smooth)
      )
      (variable
        (name Temperatur)
        (order-type quantitative)
        (data-type number)
        (properties is-density)
        (value-bound -10.2 26.2)
      )
    )
  )
)

```

```

        (precision 4 1)
        (distribution 13.4)
        (spatial-continuity high)
        (spatial-variation smooth)
    )
)
)
)
    (cardinality 1)
    (properties complete)
)
    (dimension 2)
    (coordinate-system cartesian)
    (scale 100000 to 250000)
    (acquisition-period 1997/01/01 1997/12/31)
    (acquisition representative)
)

```

#### 4.1.7 Quellen für die Datencharakterisierung

Problematisch ist in kommerziellen GIS immer noch die Akquisition und Verarbeitung der Datencharakterisierungen. Insbesondere gibt es in den meisten GIS-Datenbanken nur unzureichende Vorkehrungen für umfangreiche Datencharakterisierungen. Die Situation wird sich in zukünftigen GIS-Generationen jedoch mit Sicherheit verbessern. Ein wichtiges Indiz hierfür ist der Metadatenstandard des Federal Geographic Data Committee für die USA. Alle US-Behörden, die Geodaten erfassen, müssen seit 1995 die zugehörigen Metadaten in diesem Standard miteinfassen. Der Standard [FGDC94] enthält schon einige Elemente der vorgestellten Datencharakterisierungen, wie das geographische Bezugssystem, die Projektion und die Datenqualität.

Die Grundelemente der Datencharakterisierung sollten unbedingt schon bei der Datenerfassung mit aufgenommen werden. Zusätzliche Information können aber vom GIS oder vom Anwender auch während der Arbeit mit den Daten erkannt und dynamisch den Datencharakteristika zugefügt werden. Dazu zählen die grundlegenden statistischen und numerischen Eigenschaften der Variablen, die das GIS bei Bedarf mit einfachen numerischen Methoden ermitteln kann. Weitere algebraische und semantische Eigenschaften können bei der Analyse und Verarbeitung der Daten im GIS automatisch erkannt und mitgeführt werden. Werden z.B. Anzahlen durch eine GIS-Operation flächenmäßig normiert, haben die Ergebnisdaten offensichtlich die Dichteigenschaft.

##### Interaktive Spezifikation und Rückkopplung

Bei existierenden GIS-Datenbanken, zu denen keine Metadaten oder Datencharakteristika vorliegen, müssen diese Informationen nachträglich erfaßt und hinzugefügt werden. Dazu empfiehlt sich die Entwicklung interaktiver Werkzeuge, mit denen der Anwender diese Charakteristika spezifizieren kann.

Eine zentrale Aufgabe der Visualisierung ist natürlich auch die Unterstützung der explorativen Datenanalyse. Diese Analyse hat selbst den Hauptzweck, bisher unbekannte Ei-

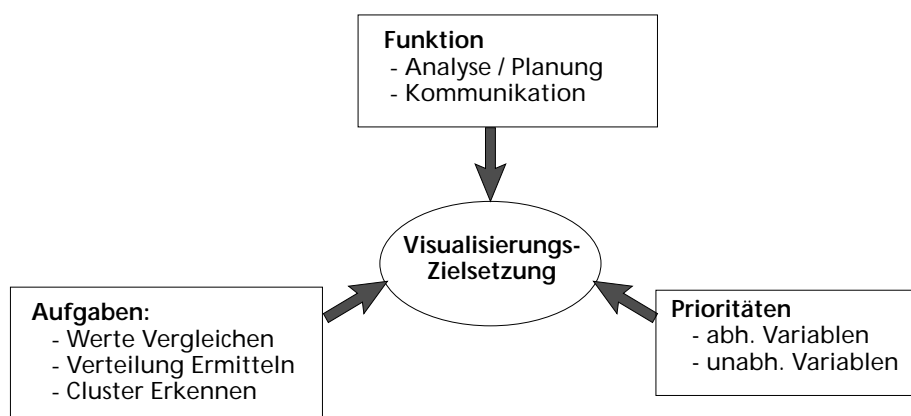
genschaften des Datensatzes zu erkennen oder hypothetische Eigenschaften zu belegen. Mit der Analyse kann insbesondere Wissen über Strukturen, Zusammenhänge und Verteilungen der Variablen erlangt werden. Dieses Wissen, auch ungesichertes Wissen in Form von Hypothesen, kann zur weiteren Charakterisierung des Datensatzes vom System festgehalten und für die nächsten Visualisierungs- und Analyseschritte nutzbringend eingesetzt werden. Eine solche Art der Rückkopplung ist in der explorativen Datenanalyse üblich.

## 4.2 Visualisierungszielsetzung

Auch das Charakterisierungsschema für GIS-Visualisierungsziele läßt sich auf einige Vorarbeiten in verwandten Gebieten aufbauen, die in der wissenschaftlichen Literatur dokumentiert sind. Wichtige Referenzen sind hier das Schema von Roth und Mattis [RoMa90], sowie das Schema von Wehrend [WeLe90]. Beide Ansätze wurden schon in Abschnitt 3.3.2 zusammengefaßt: sie haben andere Schwerpunkte, sind aber beide zu einem gewissen Grad auf das Anwendungsfeld GIS übertragbar.

Es ist jedoch notwendig, diese Schemas um diejenigen Aspekte der Visualisierungszielsetzung zu ergänzen, die spezifisch für GIS sind. Einige Vorarbeiten in dieser Richtung wurden von Turk geleistet [Turk94], seine Ergebnisse bewegen sich jedoch auf einem relativ allgemeinen Niveau, konkrete Anhaltspunkte werden kaum gegeben. Eine weitere, nützliche Quelle ist auch hier MacEachren [MacE94a].

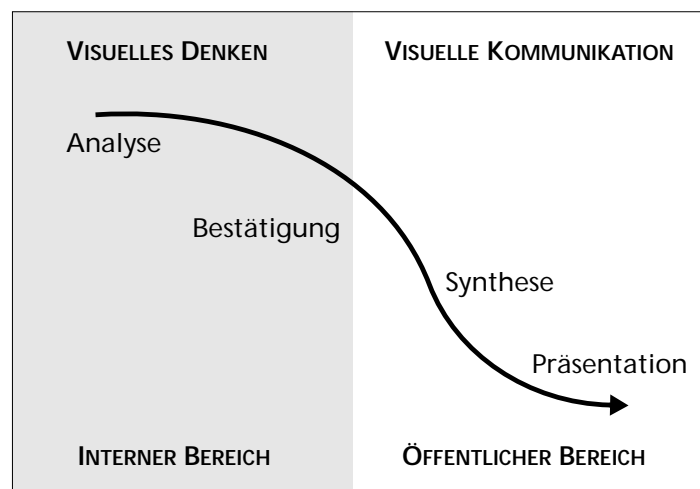
Abb. 4-6 zeigt die wichtigsten Aspekte der Visualisierung im Überblick. Auf einem relativ abstrakten Niveau angesiedelt ist die Funktion der Visualisierung. Konkrete Aussagen benennen beispielsweise die Aufgabe, die eine Visualisierung unterstützen soll und Darstellungsprioritäten bei den Variablen.



**Abb. 4-6:** Aspekte der Visualisierungszielsetzung

### 4.2.1 Funktion der Visualisierung

Bevor man sich für eine Visualisierungstechnik entscheidet, sollte man sich zunächst die Frage stellen, welche Funktion die Visualisierung erfüllen soll. Ist sie vor allem für den persönlichen Gebrauch bestimmt oder soll sie veröffentlicht bzw. einem breiten Zuschauerkreis vorgestellt werden? Dient sie eher dem Erkennen von Zusammenhängen in einem Datensatz, der Erforschung verschiedener Hypothesen in einer Planungsanwendung oder der Dokumentation von Resultaten. Folgende vier Stereotypen geben die verschiedenen Funktionen einer Visualisierung als Phasen in einem Prozeß von der Analyse bis zur Präsentation wieder:



**Abb. 4-7:** Funktion der Visualisierung, nach [MacE94a]

- *Analyse:* Erforschung eines Datensatzes und Formulieren von Hypothesen über Eigenschaften und Beziehungen der Daten. In dieser Phase werden die Visualisierungen meistens temporär erzeugt und sind für den persönlichen Gebrauch bestimmt. Die Visualisierungstechniken müssen deshalb auch für den Benutzer optimiert sein, d.h. seinen perzeptiven Fähigkeiten und auch Konventionen seiner Disziplin entsprechen. Visualisierungen sollen hier vor allem die Details zeigen und es sollen mehrere unterschiedliche Sichten des Datensatzes gleichzeitig erzeugt werden, um verschiedene, auch unvermutete Aspekte entdecken zu können.
- *Bestätigung:* Überprüfen und Bestätigen bzw. Widerlegen von Hypothesen. Hier werden vor allem Visualisierungen gefordert, die Widersprüche zu den Hypothesen oder Anomalien aufzeigen. Die Visualisierungen sind ebenfalls temporär und vor allem für den internen Gebrauch bestimmt.



- *Synthese:* Verschiedene Aspekte und Interpretationen werden zu einer Gesamt-sicht aus einer allgemeineren Perspektive zusammengebracht. Visualisierungen helfen hier sowohl bei der Integration als auch beim Wechsel der Perspektive weg von den Details und hin zum Überblick. Sie unterstützen den Anwender bei der Entscheidung, was in der Übersicht weggelassen werden kann und welche Aspekte betont werden sollen. Visualisierungen werden hier mittelfristig und für eine lokale Zielgruppe erstellt. Die Visualisierungstechniken müssen von den Experten der jeweiligen Disziplin anerkannt sein.
- *Präsentation:* In dieser Phase dienen die Visualisierungen dazu, eine Entscheidung zu präsentieren und dabei die Zuschauer zu überzeugen. Visualisierungen werden eher langfristig und für eine große Zielgruppe erstellt. Die Visualisierungstechniken müssen allgemeinverständlich sein und der Visualisierungsaufwand darf hoch sein.

#### 4.2.2 Aufgabe der Visualisierung

In jeder dieser Phasen verfolgt der Anwender mit der Visualisierung im engeren Sinne auch ganz konkrete Aufgaben. Die Visualisierungen dienen dazu, den mentalen Erkenntnisprozeß zu stimulieren und Interpretationen des Datensatzes zu erleichtern. Datensätze können aber auf vielfältigste Weise interpretiert werden oder, anders formuliert, es können unterschiedliche und auch mehrere Formen der Interpretation erwünscht sein. Der Visualisierungsentwurf soll diese verschiedenen Interpretationsformen mit einer geeigneten Visualisierung vereinfachen.

Eine naheliegende Frage ist nun: Auf welche Arten sollen Datensätze in einem GIS typischerweise interpretiert werden? Grundsätzlich sind die Interpretationsziele für Geodaten ähnlich denen für Daten anderer Disziplinen. Man will Vorkommen, Strukturen und Verteilungen von Datenwerten in Abhängigkeit von den zugrundeliegenden Dimensionen wie Raum, Zeit und anderen Variationsbereichen erkennen. Während in anderen Disziplinen die räumliche Dimension aber nur ein (wichtiger) Spezialfall ist, tritt sie bei GIS-Anwendungen vollständig in den Vordergrund. Es geht nicht nur darum, Phänomene in Abhängigkeit von ihrer absoluten Lage im Raum zu erkennen oder zu dokumentieren, sondern auch um ihre relative Lage und Entfernung zu geographischen Objekten. Insbesondere bei Planungsanwendungen ist das häufigste Visualisierungsziel, die relative Lage von Phänomenen bezüglich punktförmiger (z.B. Standorte), linienförmiger (z.B. Verkehrsnetze) oder flächenförmiger Objekte (z.B. Schutzgebiete) zu erkennen.

In der folgenden Liste sind eine ganze Reihe typischer (Interpretations-)Aufgaben zusammengestellt, wie sie für GIS-Visualisierungen auftreten. Sie reichen von so elementaren Aufgaben wie dem Ablesen von Werten bis hin zu komplexen Aufgaben wie dem Erkennen von Trends oder Clustern im räumlichen Verlauf einer Variablen. Diese Liste deckt ein breites Spektrum von Aufgaben ab, kann aber keinen Anspruch auf Vollständigkeit erheben. Je nach GIS-Anwendungsbereich können noch andere, speziellere Aufgaben dazukommen.

- *Werte Ablesen:* Alle Werte eines Datensatzes sollen möglichst exakt abgelesen werden können. Hier sind Techniken, die den Datensatz abstrahieren, z.B. durch Klassifikation, nicht geeignet. Am besten geeignet ist eine tabellarische Darstellung, sofern sie sich in übersichtlicher Form realisieren läßt. Werden visuelle Variablen eingesetzt, dann muß sichergestellt sein, daß sie die Unterscheidung von genügend vielen Werten zulassen, wie das z.B. die visuelle Variable „Länge“ für quantitative Daten zuläßt.
- *Werte Finden:* Das Vorkommen bestimmter Werte, Wertebereiche oder Wertekombinationen soll erkannt werden. Ein wichtiger Spezialfall ist das Auffinden von Minima und Maxima ordinaler oder quantitativer Datensätze. In der Visualisierung müssen die gesuchten Bereiche deutlich in den Vordergrund, die anderen Bereiche in den Hintergrund treten.
- *Werte einer Variablen Vergleichen:* Es sollen verschiedene Werte einer Variablen paarweise verglichen werden. Insbesondere bei quantitativen Daten muß die verwendete visuelle Variable einen schnellen aber exakten Vergleich von Werten ermöglichen, auch wenn die entsprechenden Symbole in der Visualisierung nicht direkt benachbart sind.
- *Werte verschiedener Variablen Vergleichen:* Für jede Kombination der unabhängigen Variablen sollen jeweils die Werte mehrerer abhängiger Variablen untereinander verglichen werden können. Handelt es sich z.B. um eine unabhängige, geographische Variable und zwei abhängige, quantitative Variablen, muß an jeder Position im Raum der Vergleich der entsprechenden Werte der zwei quantitativen Variablen möglich sein.
- *Cluster Erkennen:* Es sollen Cluster und Strukturen im Auftreten einer Variablen bezüglich ihrer Lage erkannt werden. Wichtige Fälle sind z.B. das Auftreten von Häufungen, örtlichen Schwankungen und regelmäßigen Mustern. Das Aufzeigen von Details durch die Visualisierungstechnik ist hier entscheidend. Empirische Untersuchungen zeigen, daß Choroplethenkarten hier am besten geeignet sind [LeBe95].
- *Trend Erkennen:* Es soll der globale Trend bzw. die Verteilung einer Variablen erkannt werden. Details sollen bei der Visualisierung zurücktreten und ein Gesamtüberblick soll ermöglicht werden. Beispiele solcher globaler Trends sind das „Nord-Süd-Gefälle“ und Symmetrien in der räumlichen Verteilung.
- *Korrelation Erkennen:* Es sollen geographische Korrelationen mehrerer Variablen erkannt werden [Monm90]. Wichtige Fragestellungen sind: Stimmen die Verteilungen der Variablen überein oder sind sie gegenläufig? Gibt es Abhängigkeiten zwischen den Variablen? Hier müssen im allgemeinen auch Verfahren der bivariaten und multivariaten Statistik [Witt91] eingesetzt und ihre Ergebnisse visualisiert werden.

### 4.2.3 Prioritäten

Ist die Zielsetzung der Visualisierung vielschichtig, d.h. treten mehrere unabhängige Variablen, mehrere abhängige Variablen oder mehrfache Aufgaben auf, ist die Beachtung von Prioritäten erforderlich:

- *Prioritäten bei Aufgaben:* Oft müssen mehrere Aufgaben mit unterschiedlichen Prioritäten gleichzeitig erfüllt werden, z. B. sollen gleichzeitig örtliche Strukturen und die globale Verteilung porträtiert werden. Dann muß die Visualisierungsmethode alle Aufgaben unter Beachtung der jeweiligen Prioritäten gleichzeitig optimieren.
- *Prioritäten bei unabhängigen Variablen:* Angenommen, der Datensatz ist eine Funktion mit mehreren unabhängigen Variablen und die Aufgabe ist ein Wertevergleich. Dann soll die Visualisierung den Wertevergleich zuerst bezüglich der unabhängigen Variablen mit der höchsten Priorität ermöglichen, erst dann den Vergleich bezüglich anderer unabhängiger Variablen. Soll z.B. eine abhängige Variable bezüglich Position und Zeit verglichen werden, können unterschiedliche Visualisierungstechniken notwendig sein, je nachdem, ob der Vergleich bezüglich Position oder der Vergleich entlang der Zeitachse Priorität hat.
- *Prioritäten bei abhängigen Variablen:* Sollen gleichzeitig mehrere abhängige Variablen visualisiert werden, können diesen auch Prioritäten zugeordnet sein. Ist bekannt, daß eine Variable *A* höhere Priorität als *B* hat, kann für *A* eine effektivere visuelle Variable als für *B* eingesetzt werden.

### 4.2.4 Ein formales Charakterisierungsschema

Die formale Charakterisierung einer Visualisierungszielsetzung beschreibt die Funktion der Visualisierung und eine oder mehrere Visualisierungsaufgaben:

$$\begin{aligned} \text{Objectives} &= ( \text{objectives} \\ &\quad ( \text{function } \{ \text{exploration} \mid \text{confirmation} \mid \text{synthesis} \\ &\quad \quad \mid \text{presentation} \} ) \\ &\quad \text{Task}^+ \\ & ) \end{aligned}$$

Die Aufgaben erhalten jeweils ein Gewicht und damit auch eine Reihenfolge der Priorität. Es soll zusätzlich die Forderung gelten, daß die Summe aller Gewichte einer Visualisierungszielsetzung immer 100% beträgt, das Gewicht also normiert ist.

$$\begin{aligned} \text{Task} &= ( \text{task } \text{Weight} \{ \text{Local} \mid \text{Global} \} ) \\ \text{Weight} &= \{ 0 - 100 \} \% \end{aligned}$$

Variablenlokale Aufgaben beziehen sich auf lediglich eine abhängige Variable. Sie entsprechen den in Abschnitt 4.2.2 genannten Aufgaben *Werte Ablesen*, *Werte Finden*, *Werte (einer Variablen) Vergleichen*, *Trend Erkennen* und *Cluster Erkennen*:

*Local* = local *TaskLocal DepVarName*

*TaskLocal* = { read | find | compare | trend | cluster }

Variablenübergreifende Aufgaben beziehen sich auf mehrere abhängige Variablen. Sie entsprechen den in Abschnitt 4.2.2 genannten Aufgaben (*Werte verschiedener Variablen Vergleichen* und *Korrelation Erkennen*). Bei variablenübergreifenden Aufgaben kann zusätzlich eine Prioritätsliste der unabhängigen Variablen angegeben werden.

*Global* = global *TaskGlobal DepVarName DepVarName<sup>+</sup>*  
[ domains *IndepVarName<sup>+</sup>* ]

*TaskGlobal* = { n-compare | correlate }

Dieses Schema impliziert außerdem Prioritäten bei den abhängigen Variablen. Dazu sei das Gewicht einer abhängigen Variablen definiert als die Summe der Gewichte aller Aufgaben in denen sie referenziert werden. Das Gewicht variablenübergreifender Aufgaben soll jedoch vor der Summierung durch die Anzahl referenzierter abhängiger Variablen dividiert werden. Die abhängigen Variablen sind dann in der Reihenfolge ihres Gewichts priorisiert.

### Beispiel 4-2

Das folgende Beispiel gibt eine mögliche Visualisierungszielsetzung für den bereits definierten Klimadatensatz wieder:

```
(objectives
  (function exploration)
    (task 25% compare Niederschlag)
    (task 35% find Temperatur)
    (task 40% correlate Niederschlag Temperatur
      domains Land Monat)
  )
)
```

Als Gewicht der abhängigen Variable Niederschlag ergibt sich 45%, für die Temperatur 55%. Die Temperatur hat also höhere Priorität.

## 4.3 Visualisierungskontext

### 4.3.1 Ressourcenmodell

Unter dem Begriff Ressourcen sind alle Software-, Hardware- und Peripheriekomponenten der Umgebung zusammengefaßt, die für das Rendering von Visualisierungen eingesetzt werden. Im Ressourcenmodell werden die Fähigkeiten und Charakteristika dieser Komponenten beschrieben, damit der Visualisierungsentwurf entsprechend optimiert werden kann.

### Ausgabemedien

Als wichtigstes Kriterium der Ausgabemedien ist die Farbtüchtigkeit zu nennen. Farbe spielt bei Visualisierungen auf Monitoren eine andere Rolle als beispielsweise bei gedruckten Karten: sie ist auf Monitoren praktisch ohne Zusatzkosten und -aufwand verfügbar und kann die geringere Auflösung teilweise ausgleichen. Wesentliche Parameter einer Charakterisierung sind deswegen der Grad der Farbtüchtigkeit, der von Schwarz-Weiß über Graustufen bis hin zu Farbtabelle und Vollfarbtüchtigkeit gehen kann. Wird Farbe unterstützt, sind zusätzlich Charakteristika wie Farbgamut und Farbmodell (z.B. RGB oder CMYK) interessant, da für jedes der Modelle Visualisierungsparameter optimiert werden können. Werden Visualisierungen für verschiedene Monitortypen erzeugt, dann muß der Visualisierungsentwurf sicherstellen, daß der Farbeindruck auf allen Geräten synchronisiert ist. Dazu müssen Parameter wie die Farbwiedergabe und Gammakorrektur der Monitore im Ressourcenmodell abgelegt sein.

Weitere charakteristische Merkmale der Ausgabemedien sind die Auflösung, die meistens in DPI (dots per inch) angegeben wird und die Dimensionen, d.h. der Bereich der für die Darstellung der Visualisierung verfügbar ist. Können mit dem Medium Animationssequenzen wiedergegeben werden, wie z.B. mit Monitoren oder Videorecordern, erweitert dies die Möglichkeiten der Visualisierung zeitvarianter Datensätze beträchtlich.

Für viele der dreidimensionalen Visualisierungs- und Simulationstechniken ist eine spezielle Graphikworkstation zur Darstellung erforderlich. Workstations unterscheiden sich beträchtlich in ihren Fähigkeiten bezüglich Echtzeit-Rendering von dreidimensionalen Szenen, Texturierung, Sound-Erzeugung, etc. Bei einigen Visualisierungstechniken ist eine hohe Frame-Rate erforderlich, damit der Betrachter den Eindruck von Echtzeit erhält. Ein charakteristisches Merkmal zur Beschreibung von Graphikworkstations ist deswegen auch die Renderinggeschwindigkeit, die in Polygonen pro Sekunde gemessen wird.

### Visualisierungssoftware und Interaktion

Wird eine Visualisierung nicht auf einem statischen, sondern auf einem dynamischen Ausgabemedium gerendert, sind verschiedene Stufen der Interaktion mit der Visualisierung möglich. Diese Interaktionsmöglichkeiten sind insbesondere in der explorativen Datenanalyse wichtig. Durch Vergrößerung und Verkleinerung des dargestellten Ausschnitts läßt sich mit einer entsprechend intelligenten Software das Wechselspiel zwischen Detail und Überblick sehr dynamisch realisieren. Manche Visualisierungstechniken erfordern einen noch höheren Grad an Interaktion, dazu gehören insbesondere die Techniken bei denen die Visualisierung zugleich auch Simulation ist. Ein Beispiel ist das *Particle Tracing*, bei dem eine Partikelwolke interaktiv positioniert werden muß.

Die Interaktionsfähigkeit wird vor allem durch die Funktionalität der Visualisierungssoftware bestimmt. Sie kann in verschiedenen, aufeinander aufbauenden Stufen gemessen werden:

- *Stufe 1:* Die Ansicht der Visualisierung kann variiert werden. Wesentliche Funktionen sind Zoom-In, Zoom-Out, Value-Query, Animationssequenz abspielen, 3D-Viewing-Parameter ändern.

- *Stufe 2:* Die Parameter der Visualisierungstechniken können variiert werden, z.B. die Farbpalette, die Symbolisierung oder die Partikelpositionen.
- *Stufe 3:* Ein der Visualisierung vorausgehender Simulationsprozeß läßt sich aus der Darstellung heraus steuern. Diese Interaktionsform wird auch *Steering* genannt.

Ein weiteres charakteristisches Merkmal der Visualisierungssoftware sind die Visualisierungstechniken, die sie unterstützt.

### 4.3.2 Benutzermodell

Informationen über Benutzer oder Zielgruppen können eingesetzt werden, um die Visualisierung speziellen, persönlichen Anforderungen anzupassen. Diese Personalisierung ist allerdings nur dann möglich und wünschenswert, wenn die Visualisierung lediglich für den persönlichen Gebrauch oder für einen eingeschränkten Nutzerkreis vorgesehen ist

Mit einem Modell zur symbolischen Beschreibung von Benutzern läßt sich überprüfen, ob spezifisches Wissen vorhanden oder nicht vorhanden ist, das zum Verständnis einer speziellen Visualisierungstechnik erforderlich ist. Benutzermodelle können sogar Informationen über spezifische Charakteristika von Benutzern enthalten, die eine Modifikation der Standardtechniken erfordern. Hier ist in z.B. die Farbenblindheit zu nennen, von der immerhin ca. 8% der Männer und ca. 0.5% der Frauen betroffen sind.

Benutzermodelle haben eine gewisse Tradition in der KI-Forschung und -Anwendung und die Literatur zu diesem Thema ist entsprechend umfangreich. Allen hat eine Taxonomie aufgestellt, die die drei wesentlichen Dimensionen von Benutzermodellen aufzeigt [Alle90]. Jedes Benutzermodell läßt sich durch eine Position bzw. ein Intervall in den folgenden drei Dimensionen charakterisieren:

- *Canonical vs. individual models:* Enthält das System ein Modell für eine Reihe von kanonischen Benutzern (d.h. Stereotypen) oder für jeden individuellen Benutzer?
- *Explicit vs. implicit models:* Werden die Benutzermodelle beim Systementwurf explizit beschrieben oder werden sie aus dem Verhalten der Benutzer hergeleitet?
- *Long-term vs. short-term models:* Beschreiben die Modelle eher Langzeitcharakteristika (z.B. Fähigkeiten und Kenntnisse) oder eher Kurzzeitcharakteristika (z.B. das Problem, das der Benutzer gerade löst)?

Für den Visualisierungsentwurf sind sicher verschiedene Arten von Benutzermodellen sinnvoll und fast alle Kombinationen aus dieser Taxonomie denkbar. Eine ausführliche Betrachtung des Themas Benutzermodelle in GIS und bei der GIS-Visualisierung würde aber den Rahmen dieser Arbeit übersteigen und könnte sicher eine eigenständige wissenschaftliche Arbeit füllen. Deswegen soll an dieser Stelle das Thema Benutzermodelle in

GIS nur exemplarisch und ausschließlich vom Standpunkt der Visualisierung aus betrachtet werden.

Ein geeignetes Benutzermodell für den Visualisierungsentwurf ließe sich mit dieser Taxonomie wie folgt charakterisieren: Jeder Benutzer ist individuell modelliert. Die Modellierung erfolgt explizit, indem das System die Parameter in einem Interview erfragt. Es werden vor allem die Fähigkeiten in der visuellen Wahrnehmung und die Kenntnisse in der Visualisierung modelliert, also Langzeitcharakteristika.

Im folgenden werden einige wesentliche Parameter eines solchen Modells vorgeschlagen. Für spezielle Anwendungen sind jedoch auch andere Arten der Benutzermodellierung denkbar, die folglich andere Positionen in der Taxonomie einnehmen würden. So kann zum Beispiel das Verhalten des Benutzers beim Auswählen und Modifizieren von Visualisierungsentwürfen beobachtet werden und damit auf Präferenzen geschlossen werden.

### **Physiologie und Perzeption**

Zur Modellierung von Wahrnehmung und Motorik läßt sich das Modell von Domik und Gutkauf verwenden, das bereits in Abschnitt 3.3.3 vorgestellt wurde. Die Parameter dieses Modells sind:

- Farbwahrnehmung
- Farbklassifizierung
- Farbgedächtnis
- räumliche Vorstellung
- Feinmotorik

### **Kenntnisstand**

Für die Modellierung des Kenntnisstandes läßt sich kein allgemeingültiges Schema definieren. Mit welchen konkreten Parametern der Kenntnisstand am geeignetsten repräsentiert wird, hängt sehr stark von der Anwendung ab.

Allgemeine Kenntnisse auf dem Gebiet der Visualisierung und Kartographie lassen sich zum Beispiel durch eine Auflistung der dem Benutzer bekannten Visualisierungstechniken nur annähernd festhalten. Das System muß zusätzlich wissen, welche grundlegenden Gesetzmäßigkeiten der graphischen Kommunikation der Benutzer gelernt und verstanden hat. Online-Tutorials sind an dieser Stelle eine Möglichkeit, solche Kenntnisbereiche zu strukturieren und den Kenntnisstand des Benutzers zu modellieren.

Ein weiterer Aspekt des Kenntnisstandes ist eine Ausbildung in einer Spezialdisziplin, die spezifische Visualisierungstechniken und -prinzipien einschließt. Schließlich ist noch die Nutzungshäufigkeit und das Nutzungsniveau der Visualisierungsumgebung selbst zu nennen. Stereotypen wie *novice*, *casual*, *standard* und *power user* bieten sich für die Modellierung an.

### Präferenzen

Für den Visualisierungsentwurf sind u.a. folgende Präferenzen des Benutzers relevant:

- persönliche Gestaltungsvorlieben wie Vordergrund- und Hintergrundfarbe sowie typographische Attribute
- persönliches Nutzungsprofil der Umgebung und der Visualisierungssoftware
- bevorzugte Visualisierungstechniken

### 4.3.3 Ein formales Kontextmodell

Wie sich aus der Diskussion schon ergeben hat, ist das Kontextmodell der am stärksten applikations- und umgebungsabhängige Faktor beim Visualisierungsentwurf. Deshalb soll hier nur der anwendungsübergreifende Teil eines Kontextmodells formal spezifiziert werden. Aus dem gleichen Grund wird zur Modellierung von Kenntnissen und Präferenzen nur mit relativ allgemeinen Stereotypen gearbeitet.

Das formale Kontextmodell besteht aus einem Ressourcen- und einem Benutzermodell:

```
Context      = ( context
                  ResourceModel
                  UserModel
                )
```

### Ressourcen-Modell

Das Ressourcen-Modell beschreibt Ausgabemedium und Interaktionsmöglichkeiten. Für das Ausgabemedium werden die Farbtüchtigkeit und das verwendete Farbmodell charakterisiert. Zur Modellierung der Auflösung dienen Stereotypen: Video (sehr niedrige Auflösung, z.B. 50 DPI), Monitor (z.B. 80 DPI), Laserdrucker (300 bzw. 600 DPI) und Photographische Reproduktion bzw. Druck mit mindestens 1200 DPI. Die Dimensionen des Ausgabemediums sind in der Einheit mm angegeben. Dynamische Möglichkeiten der Ausgabe können animierte Bildsequenzen, Rendering von 3D-Modellen und das Texturieren von Modellen sein.

Möglichkeiten der Interaktion sind Zooming, Abfragen von Datenwerten in der Visualisierung, Steuerung von Animationssequenzen, Definition von Viewing-Parametern und Modifikation der Visualisierungsparameter.

```
Ressource= ( output-medium
              ( name MediumName )
              ( color { bi-tone | gray-tone | pseudo-color | full-color } )
              [ ( color-space { RGB | CMYK } ) ]
              ( resolution { video | monitor | laser-low | laser-high
                             | photo } )
              ( dimensions Number Number )
```



```

[ ( dynamic [ animation ] [ rendering ] [ texturing ] )
  [ ( interaction [ zoom ] [ query ] [ animation ]
    [ viewing ] [ modify ] ) ]
)

```

Im folgenden Beispiel werden zwei typische Ausgabemedien in diesem Modell charakterisiert: eine Graphikworkstation und hochauflösender Laserdrucker:

```

(output-medium
  (name Workstation:SGL-Indigo)
  (color full-color)
  (color-space RGB)
  (resolution monitor)
  (dimensions 325 254)
  (dynamic animation rendering texturing)
  (interaction zoom query animation viewing modify)
)

(output-medium
  (name Laserprinter:HP-6000)
  (color bi-tone)
  (resolution laser-high)
  (dimensions 210 297)
)

```

### Benutzermodell

Das Benutzermodell umfaßt die Möglichkeit, individuelle Abweichungen in der Farbwahrnehmung festzuhalten. Bei der anomalen Farbensichtigkeit wird zwischen den wichtigsten Typen Deuteranomalie (mit ca. 5% Häufigkeit unter der männlichen Bevölkerung), Deuteropie, Protanomalie und Protanopie (mit jeweils ca. 1% Häufigkeit unter der männlichen Bevölkerung) unterschieden. Deuteranomale und deuteropie Personen verfügen über eine verminderte Unterscheidungsfähigkeit zwischen Rot- und Grüntönen. Protanomale und Protanope nehmen zusätzlich rötliche Farben dunkler war. Nur ca. 0.5% der Frauen sind von einer dieser Formen der anomalen Farbensichtigkeit betroffen. (Andere Formen werden hier wegen ihrer sehr geringen Häufigkeit, jeweils unter 0.001%, nicht berücksichtigt.) Ist dem Entwurfssystem die Anomalie in der Farbensichtigkeit eines Benutzers bekannt, so können für solche Personen besonders geeignete Farbskalen für thematischen Karten [OlBr97] und für Visualisierungen [MeGr85] verwendet werden.

Das Niveau der Systemkenntnis und Ausbildung des Benutzers läßt sich mittels dreier Stereotypen klassifizieren: Anfänger, „Standard“-Benutzer und Experte. Eine detaillierte Information über den Kenntnisstand des Benutzers wird über die Liste der von ihm erfolgreich absolvierten (oder als schon bekannt eingestuft) Kurse abgelegt. Schließlich können allgemeine Präferenzen des Benutzers abgelegt werden, die beim Aufruf der Umgebung die Standardwerte überschreiben.

```

UserModel = ( user
              ( name UserName )

```

```

[ ( color-deficiency [ protanomalous | protanopic
    deutanomalous | deutanopic ] ) ]
( system-knowledge { expert | regular | novice } )
( visualization-knowledge { expert | regular | novice } )
[ ( course-units UnitName+ ) ]
[ ( preferences KeyValuePair+ ) ]
)

KeyValuePair = ( KeyName Value )

```

#### 4.3.4 Quellen des Kontextmodells

Wie können die Parameter des Kontextmodells erfaßt werden? Für die Ressourcen läßt sich dieses Problem relativ einfach lösen. Hardware- und Softwarekomponenten ändern sich während ihrer Lebensdauer meist nur geringfügig und lassen sich formal und objektiv anhand einiger weniger Parameter beschreiben. Ihre Charakteristika müssen deshalb nur einmal, vor Beginn ihrer Inbetriebnahme in der Visualisierungsumgebung erfaßt werden.

Wesentlich komplexer ist die Erfassung der Benutzermodelle, da sie zum Teil dynamisch und schwerer formalisierbar sind. Hier sollten mehrere Erfassungsmethoden parallel verwendet werden, damit das Benutzermodell inkrementell verbessert und aktualisiert werden kann:

- *Interviews:* Das System fragt den Benutzer explizit nach einer Reihe von charakteristischen Attributen, wie z.B. seinen Gestaltungspräferenzen. Das Interview wird in der Regel vor der ersten Visualisierungssitzung durchgeführt.
- *Interaktive Tests:* Einige Parameter lassen sich mit Tests messen, beispielsweise messen Domik und Gutkauf [DoGu94] mit einem Computerspiel Abweichungen in der Perzeption, Kognition und Motorik.
- *Protokollierung:* Aussagekräftiges Verhalten des Benutzers bei der Arbeit mit dem System wird protokolliert, wie z.B. absolvierte oder abgelehnte Kurseinheiten des Online-Tutorials.
- *Beobachtung:* Präferenzen werden durch Beobachten des Benutzerverhaltens ermittelt: Welche der vom System vorgeschlagenen alternativen Visualisierungen werden am häufigsten ausgewählt? Welche vom System gewählten Parameter werden vom Benutzer nachträglich geändert?

Schlußfolgerungen und Verallgemeinerungen dürfen allerdings nur gemacht werden, wenn sich im Benutzerverhalten ein klares Schema ergibt. Werden Schlußfolgerungen über Präferenzen zu schnell gezogen und umgesetzt, wird der Benutzer durch das wechselnde Systemverhalten eher verwirrt als entlastet. Wenn Muster im Benutzerverhalten

vom System scheinbar erkannt werden, sollte es den Benutzer explizit danach fragen, z.B.:

*Ziehen Sie normalerweise für den Vergleich von quantitativen Variablen in großen Maßstabsbereichen Kreissignaturen den Choroplethen vor?*

## 4.4 Zusammenfassung und Diskussion

### Zusammenfassung der Ergebnisse

In Kapitel 4 wurde untersucht, welche Informationen über Daten und GIS-Umgebung vorhanden sein müssen, damit effektive GIS-Visualisierungen automatisch entworfen werden können. Als Ergebnis steht ein Modell, das eine formale Charakterisierung des variablen Kontextes einer Visualisierung erlaubt. Das Modell besteht aus vier Teilen:

1. *Datencharakteristika:* Neben den algebraischen und strukturellen Eigenschaften, die auch für Datensätze aus anderen Anwendungsgebieten relevant sind, spielt bei GIS-Daten der Raumbezug eine besondere Rolle. Dem wurde durch die Einführung eines geographischen Variablentyps entsprochen, der die klassische Dreiteilung (nominal, ordinal, quantitativ) ergänzt. Als charakterisierende Eigenschaften wurden u.a. Maßstab, Projektion, geographische Abdeckung und Stetigkeit sowie verschiedene Parameter der Datenqualität herausgearbeitet und formalisiert.
2. *Visualisierungszielsetzung:* Diese kann auf einer abstrakten Ebene der Visualisierungsziele charakterisiert werden, die von der explorativen Analyse bis zur Synthese und Kommunikation von Ergebnissen reichen. Es lassen sich aber auch häufig konkrete Aufgaben der Visualisierung benennen, die durch eine Struktur aus Operatoren, Variablen und Gewichten beschrieben werden.
3. *Benutzercharakteristika:* Für die Effektivität einer Visualisierung sind vor allem zwei Eigenschaften der Benutzer bestimmend: visuelle Fähigkeiten wie Farbwahrnehmung und räumliche Vorstellung, sowie die Kenntnis disziplinspezifischer Visualisierungstechniken.
4. *Ausgabemedien-Charakteristika:* Hierzu zählen Kenngrößen wie Farbtüchtigkeit, Auflösung und Dimensionen, im Falle dynamischer Ausgabemedien auch verschiedene Grade der Interaktivität.

### Diskussion

Dieses Modell führt frühere Ansätze zur Charakterisierung von Daten und Kontext für die automatisierte Visualisierung fort. Wichtige Quellen waren [RoMa90] für fundamentale Datencharakteristika, [WeLe90] für Interpretationsziele und [MacE94a] für Geodaten-spezifische Charakteristika.

Als neue und signifikante Aspekte des in dieser Arbeit entwickelten Charakterisierungsmodells sind hervorzuheben:

- die aus der Literatur bekannten Ansätze wurden für das Anwendungsgebiet Geodaten-Visualisierung erheblich erweitert
- die verschiedenen Modelle wurden in einem Gesamtmodell konsistent integriert
- basierend auf das Modell wurde eine Syntax zur formalen Beschreibung von Daten- und Kontextcharakteristika aufgestellt

Das entwickelte Charakterisierungsmodell kann über seinen engeren Zweck als Fundament einer Methodik des automatisierten GIS-Visualisierungsentwurfs auch in anderen Bereichen gewinnbringend eingesetzt werden. Viele der beschriebenen charakteristischen Attribute sind nicht nur für den Entwurf von Visualisierungen in GIS, sondern auch in anderen Anwendungsbereichen relevant, so z.B. in der Visualisierung technisch-wissenschaftlicher Datensätze und der Informationsvisualisierung. Andererseits unterstützt insbesondere das Datenmodell aber auch die Konsistenz und Effektivität anderer Operationen in GIS, wie z.B. Analyse- und Verarbeitungsoperationen und die Integration von raumbezogenen Datensätzen aus unterschiedlichen Quellen, die sogenannten *Conflation*. Es kann deswegen als Ausgangspunkt für die Entwicklung eines umfangreichen Metadatenkatalogs für Geodatenbestände dienen.

### Ausblick

Als Gegenstand weiterer Forschung bietet sich unter anderem das Modell der Visualisierungszielsetzung an. Es beschreibt in seiner bisherigen Form zwar einige zentrale Aspekte der Zielsetzung einer Visualisierung, die vor allem in der Phase der explorativen Analyse auftreten können. Es genügt aber nicht mehr zur Beschreibung der komplexen Aufgaben, die Visualisierungen in den späteren Phasen, z.B. beim Testen von Hypothesen und in der Synthese und Präsentation von Ergebnissen unterstützen sollen. Hier könnte z.B. ausgehend von den Ergebnissen Casners [Casn91] in der Aufgabenmodellierung eines eng umgrenzten Anwendungsfeldes, versucht werden, allgemeinere logische Operatoren zu benennen. Ergebnisse solcher Forschungsarbeiten dürften dem automatisierten Visualisierungsentwurf in allen Anwendungsbereichen wichtige neue Impulse geben.

# Kapitel 5

## Eine Methodik für den automatisierten GIS-Visualisierungsentwurf

Kapitel 5 leitet eine Methodik für den automatisierten Entwurf von Visualisierungen raumbezogener Sachdaten her, die als Grundlage für die Implementierung konkreter, intelligenter Visualisierungshilfen dient. Zunächst werden exemplarisch verschiedene Arten von Regeln vorgestellt, die in Lehrwerken zur thematischen Kartographie und Visualisierung graphisches Wissen beschreiben. Aus der Struktur dieser Regeln lassen sich erste Aussagen über geeignete Formen der Repräsentation in Expertensystemen treffen.

Neben einer Regelbasis zur Ausdrucksfähigkeit und Effektivität der einzelnen Visualisierungstechniken benötigt man aber auch eine geeignete Strategie zum Entwurf von Visualisierungen verschiedenster raumbezogener Datensätze. Basierend auf den grundlegenden Arbeiten von Mackinlay wird eine Entwurfsstrategie hergeleitet, die für thematische Karten und andere GIS-Visualisierungen geeignet ist. Sie setzt die Regeln zur Ausdrucksfähigkeit ein, um Teilvisualisierungen des Datensatzes zu finden und verwendet Kompositionsregeln, um sie dann zu vollständigen Visualisierungen zusammenzufügen. Eine Besonderheit der vorgestellten Entwurfsstrategie ist, daß sie nicht nur eine, sondern normalerweise eine ganze Reihe ausdrucksfähiger Visualisierungen für den Datensatz entwirft, diese bezüglich ihrer Effektivität bewertet [Jung95, Jung96a] und dem Anwender mehrere effektive Visualisierungsmöglichkeiten vorschlägt.

Die Bewertung der Effektivität basiert zum einen auf einem Fuzzy-Schema zur Modellierung und Abschätzung von Visualisierungseffektivität, das auch den Visualisierungskontext berücksichtigt. Zum anderen basiert sie auf konkreten Maßen für die Effektivität verschiedener Visualisierungstechniken für unterschiedliche Aufgaben, die aus der Literatur entnommen bzw. im Rahmen dieser Arbeit in einer Reihe psychologischer Tests ermittelt wurden. Im Anschluß an die Vorstellung der eigentlichen Entwurfsmethodik in Kapitel 5 werden in Kapitel 6 diese Tests und ihre Ergebnisse diskutiert.

## 5.1 Repräsentation graphischen Wissens

Neben den Inferenzalgorithmen waren die Wissensrepräsentationsformen immer ein zentrales Thema der Künstlichen Intelligenz (KI). Die überwiegende Mehrzahl der heute arbeitenden KI-Softwaresysteme repräsentiert prozedurales Wissen durch **IF-THEN**-Konstrukte und semantisches oder strukturelles Wissen durch Frames und Netzwerke. Kommerziell verfügbare Expertensystemshells bieten diese Repräsentationsformen standardmäßig an. Von Bedeutung ist außerdem die Wissensrepräsentation durch eine Prädikatenlogik zweiter Stufe, auf der Prolog aufbaut.

Praktisch alle wissensbasierten Systeme für den Visualisierungsentwurf verwenden Expertensystemshells und die zugehörigen Repräsentationsformen. Dafür gibt es zwei wesentliche Gründe. Zum einen ist die prozedurale **IF-THEN**-Repräsentation sehr mächtig und eignet sich auch für die direkte Wiedergabe graphischer Regeln. Im Laufe dieses Kapitels werden noch zahlreiche Beispiele solcher Regeln und ihrer Repräsentation gegeben. Zum anderen sind Expertensystemshells mittlerweile weit verbreitet, zuverlässig und ausgereift. Teilprobleme des graphischen Entwurfs lassen sich auch mit der Prädikatenlogik formulieren. Ein Beispiel dafür wurde schon in Abschnitt 3.4.2 genannt: die automatisierte Kartenbeschriftung.

### 5.1.1 Klassen graphischer Regeln

Zunächst sollen exemplarisch einige Formen graphischer Regeln untersucht werden. Dazu wurden aus der kartographischen Literatur Regeln für zwei der wichtigsten Techniken der thematischen Kartographie, Choroplethen und Kreissignaturen extrahiert. Es wurden auch allgemeinere Regeln über die Verwendung von Farbe und anderen visuellen Variablen betrachtet. Quellen waren zwei Lehrwerke zur thematischen Kartographie, die von Borden Dent [Dent96] und Judith Tyner [Tyne92] stammen.

Die Regeln lassen sich zunächst grob in drei thematische Klassen einteilen. Diese Einteilung spiegelt erneut die drei Aspekte guten graphischen Entwurfs wieder:

- *Regeln zur Ausdrucksfähigkeit:* Unter welchen Bedingungen kann eine Technik für einen Datensatz verwendet werden? Wann muß die Verwendung der Technik vermieden werden?
- *Regeln zur Effektivität:* Welche Technik ist unter gegebenen Bedingungen die effektivere bzw. die effektivste? Welche Bedingungen zeigen Fälle an, in denen eine Technik weniger effektiv oder besonders effektiv ist?
- *Regeln zur visuellen Ausgestaltung und Parameterwahl:* Welche Farben oder visuellen Konstanten sollen unter gegebenen Bedingungen und für eine bestimmte Technik gewählt werden? Wie harmonisieren verschiedene Farben oder Signaturen miteinander?

### 5.1.2 Ausdrucksfähigkeit

Regeln zur Ausdrucksfähigkeit sind gewöhnlich von einer einfachen Form: eine Reihe von Bedingungen muß erfüllt sein, damit eine Technik ausdrucksfähig ist. Unter den Bedingungen finden sich naturgemäß vor allem Restriktionen bezüglich des Datensatzes, seines Ordnungstyps, seiner Verteilung, etc. Die folgenden zwei Beispiele für solche Regeln geben Aufschluß über die Verwendung von Choroplethen:

If data have been collected for statistical areas instead of at sample points, or if the distribution is discontinuous, the statistical surface will be step-like and one of the choropleth mapping techniques will be suitable. [Tyne92, S. 170]

Unless the enumeration areas are all the same size, densities, ratios, percentages, or other derived values must be used, rather than absolute values, for the map to be meaningful. [Tyne92, S. 171]

Jede dieser kartographischen Regeln nennt also notwendige, aber im allgemeinen nicht hinreichende Bedingungen für die Ausdrucksfähigkeit einer Technik. Eine vollständige Regel über die Ausdrucksfähigkeit einer Technik ergibt sich somit aus der Konjunktion dieser Bedingungen:

**IF** *Bedingung*(*Datensatz*)  $\wedge$  *Bedingung*(*Datensatz*)  $\wedge$  ...  
**THEN** *Technik*(*Datensatz*) ist ausdrucksfähig.

### 5.1.3 Effektivität

Regeln zur Effektivität treten in verschiedenen Formen auf. In der einfachsten Form lautet die Regel: **IF** *Bedingung* **THEN** *Technik* ist effektiv. Die Bedingungen beziehen sich hier meist auf die Visualisierungszielsetzung und auf die Charakteristika des Datensatzes. Ein Beispiel dafür ist die folgende Regel:

A choropleth map should be used when the purpose is to compare enumeration areas. [Tyne92, S. 171]

Häufig sind die Aussagen zur Effektivität jedoch differenzierter und ordnen verschiedenen Visualisierungszielsetzungen verschiedene Grade an Effektivität oder auch Ineffektivität zu. Die folgenden beiden Regeln etwa bewerten die Effektivität der Choroplethentechnik bezüglich verschiedener Visualisierungsaufgaben:

Map readers use choropleth maps in three ways: to ascertain an actual value associated with a geographical area, to obtain a sense of the overall geographic pattern of the mapped variable with attention to individual values, and to compare one choropleth map pattern to another. [Dent96, 123]

Limitations of the choropleth technique are (1) that the uniform shading of enumeration areas can give the impression of uniform values, and (2) that exact values cannot be determined. [Tyne92, S. 171]

Andere Arten von Effektivitätsregeln lassen sich wegen der Verwendung unscharfer Bedingungen weniger leicht in das einfache **IF-THEN** Schema einordnen. Eine Regel über die Effektivität von Kreissignaturen bezieht sich beispielsweise auf „geringe“ Schwankungen im Datensatz:

Proportional point symbol mapping should not be selected when data variation is small or when the nature of the classing method renders a map of homogenous appearance. [Dent96]

Regeln können auch eine weitere Dimension der Unschärfe enthalten, wenn sie nicht allgemein anerkannt sind oder zwischen verschiedenen kartographischen Schulen Uneinigkeit über die Effektivität einer Technik unter gegebenen Bedingungen besteht. Die folgenden beiden Regeln illustrieren diese Art der Unschärfe am Beispiel der Kreissignaturen:

It appears that range-grading proportional symbols may be a better choice than apparent scaling. Range-grading of proportional point symbols raises the issue of class interval determination. The most applicable method appears to be the optimization approach. We recommend that perhaps only four or five classes be used. [Dent96]

Some cartographers feel that range-graded circles may be the best proportional symbol. [Tyne92, S. 153]

Neben denjenigen Regeln, die die Effektivität von individuellen Techniken bewerten, gibt es natürlich auch Regeln über die Effektivität von komplexeren Visualisierungen, die mehrere Techniken vereinen:

It is not uncommon for two proportional forms to illustrate two distributions on the same map. This is to be discouraged because it introduces too much complexity on the map and can undermine the communicative effort. A better solution would be to develop two separate maps. [Dent96]

Diese Beispiele zeigen, daß die Anforderungen an die Repräsentationsform für Effektivitätsregeln vielschichtiger sind als bei Regeln zur Ausdrucksfähigkeit. Die folgende Aufzählung faßt die wichtigsten Anforderungen zusammen:

- Bedingungen zur Effektivität können sich auf verschiedene Aspekte, wie Datencharakteristika, Zielsetzung und Kontext der Visualisierung beziehen.
- Regeln weisen einer Technik unter verschiedenen Bedingungen verschiedene Grade der Effektivität zu. Zwischen den beiden Extremen „effektiv“ und „nicht effektiv“ muß es demnach eine Reihe von Abstufungen geben.
- Unsicherheiten und Unschärfen können in den Regeln enthalten sein.
- Da sehr verschiedene Aspekte (Datencharakteristika, Zielsetzung und Kontext) sich auf die Effektivität auswirken, kann die Effektivitätsbewertung insgesamt widersprüchlich sein, d.h. eine Technik unter gegebenen Bedingungen kann sowohl effektive als auch ineffektive Aspekte vereinen.

Insgesamt läßt sich jetzt schon sagen, daß zur Bewertung der Effektivität mehrere, sich ergänzende Regeln existieren müssen. Die Effektivität selbst läßt sich weder durch eine



Boole'sche Größe noch durch eine lineare Skala hinreichend repräsentieren. Um Unsicherheiten und Widersprüche auszudrücken bieten sich am ehesten Fuzzy-Mengen an, wie in Abschnitt 5.7.2 noch gezeigt wird.

#### 5.1.4 Ausgestaltung und Parameterwahl

Diese dritte Klasse von Regeln unterscheidet sich von den beiden anderen dadurch, daß sie erst nach der Entscheidung für eine Technik zur Anwendung kommt. Diese Regeln bestimmen dann, wie eine Visualisierung visuell ausgestaltet und wie die Parameter der enthaltenen Techniken gewählt werden müssen, damit eine hohe Effektivität erreicht wird.

Manche dieser Regeln sind sehr allgemein und insbesondere auch unabhängig von Datensatz, Zielsetzung und Kontext. Sie geben allgemeingültige Schranken für die Wahl eines Parameters an, oder legen diesen Parameter sogar fest. Die folgenden beiden Regeln sind allgemeine Aussagen über die Klassifizierung von quantitativen Datensätzen für thematische Karten:

Arbitrary categories, such as those based on round numbers or a division of the data in equal-sized categories, generally are not desirable. [Tyne92, S. 174]

Normally, no fewer than four and no more than ten categories are used. Four to six steps is the most common number of categories. [Tyne92, S. 175]

Andere Regeln legen bestimmte Parameter abhängig von Bedingungen an den Kontext oder den Datensatz fest:

On a black and white display, the categories are represented by shades of gray; for color maps, different value steps of a hue are used. The accepted convention is that high values (light tones) represent low numbers and low values (dark tones) represent high numbers. [Tyne92, S. 175, 184]

Color conventions include blue for water, green for vegetation and red for cities. [Tyne92, S. 258]

Der dritte und komplexeste Regeltypus beinhaltet wieder eine Bewertung der Effektivität. Hier werden entweder verschiedene Möglichkeiten der Parameterwahl bezüglich ihrer Effektivität verglichen, oder Wege vorgeschlagen, bestimmte Aspekte der Effektivität einer Visualisierung zu erhöhen.

Although *tones* can be used to represent differences in kind, it is not a good practice. The more common convention is for differences in tonal value to symbolize differences in degree or amount. Normally, light tones are used to represent small amounts and dark tones represent larger amounts. [Tyne92, S. 135]

Color has particular potential for increasing the effectiveness of multivariate representations, but only if it is matched logically to data characteristics. At a simple level, one can depict a set of numerical values that spans some critical value as bidirectional (or diverging from this critical value). With this conceptualization, a hue difference makes sense as a way to distinguish

the two ends of the scale and a value difference within each hue can be used to depict the numerical ranges. [MacE94a, S. 91]

Wegen ihrer heterogenen Erscheinung kann für die Regeln zur Parameterwahl keine allgemeine Form festgelegt werden. Allgemeingültige Ausgestaltungsregeln können direkt in eine Einschränkung oder Festlegung von Parametern umgesetzt werden und müssen in die Wissensbasis nicht aufgenommen werden.

## 5.2 Eine Strategie für den GIS-Visualisierungsentwurf

Neben Regeln zur Ausdrucksfähigkeit und Effektivität von Visualisierungstechniken benötigt ein automatisiertes Visualisierungssystem auch eine Entwurfsstrategie. Diese Strategie beschreibt im Detail, wie für jeden Datenfall eine regelkonforme Visualisierung gefunden werden kann. Damit eine Implementierung in einem Softwaresystem möglich ist, muß sich die Entwurfsstrategie auch als Algorithmus formulieren lassen.

In der Literatur werden eine ganze Reihe verschiedener Strategien für die Lösung von Entwurfsproblemen genannt [FeMM92]. Drei dieser Strategien eignen sich insbesondere auch für den Entwurf graphischer Darstellungen:

- *Fall-basierter Entwurf:* Ein neues Entwurfsproblem wird auf eine Anzahl bekannter Probleme (Fälle) zurückgeführt. Lösungen dieser Fälle werden für die Lösung des neuen Problems verwendet und gegebenenfalls modifiziert.
- *Entwurf durch Komposition:* Der Entwurf wird aus Teilstücken zusammengesetzt. Die Regeln für die Komposition können einfach, wie bei der Überlagerung transparenter Folien, aber auch komplex, wie beim Layout strukturierter Dokumente sein.
- *Entwurf als Optimierung:* Der Entwurf läßt sich als lineares Optimierungsproblem formulieren und kann folglich mit einem der bekannten Optimierungsalgorithmen (z.B. Simplex-Algorithmus, lineare Programmierung) gelöst werden.

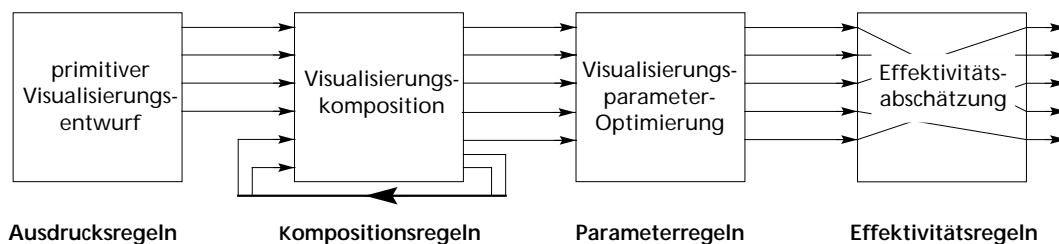
Ein Fall-basierter Entwurf ist praktisch nur durchführbar, wenn sich das Problem auf eine überschaubare Zahl von Fällen zurückführen läßt. Damit ist dieser Ansatz für den Visualisierungsentwurf ungeeignet, da sich durch die Variationsmöglichkeiten von Datensatz, Zielsetzung und Kontext eine sehr große Zahl verschiedener Fälle ergibt. Auch läßt sich der Visualisierungsentwurf nur in Teilaspekten als lineares Optimierungsproblem definieren. Insbesondere die komplexen Gesetzmäßigkeiten der Ausdrucksfähigkeit und Effektivität können nicht auf wenige zu optimierende Variablen reduziert werden.

Der Entwurf durch Komposition hat sich deswegen seit APT als Standardstrategie für den Visualisierungsentwurf durchgesetzt. Nur mit Kompositionsoperatoren läßt sich ein ausreichend großer Entwurfsraum für Visualisierungen garantieren. Der Entwurf durch Komposition kommt dabei auch der Arbeitsweise moderner Softwaresysteme für die technisch-wissenschaftliche Visualisierung nahe. Solche Systeme, wie z.B. APE und AVS, bieten dem Anwender eine Reihe von Visualisierungstechniken für Datensätze an.

Dazu gehören Operationen, mit denen Visualisierungen modifiziert oder kombiniert werden können. Der Anwender entwirft komplexe Visualisierungen in einer graphischen Benutzerschnittstelle durch Verschieben und Verknüpfen ikonischer Repräsentationen der Visualisierungstechniken und Operationen. In einem entsprechenden GIS-Visualisierungssystem zählen zu den Techniken beispielsweise die thematischen Karten und das 3D-Rendering statistischer Oberflächen. Zu den Kompositionsoperationen zählen die kartographische Überlagerung und die Animation.

### 5.3 Die Entwurfsmethodik im Überblick

Nach diesen grundsätzlichen Betrachtungen zur Repräsentation graphischer Regeln und zur Strategie des automatischen Entwurfs soll die Entwurfsmethodik für GIS-Visualisierungen nun im Überblick dargestellt werden. Die Methodik gliedert sich in vier Hauptphasen: primitiver Entwurf, Komposition, Parameterselektion bzw. -optimierung und Effektivitätsabschätzung. Alle vier Phasen der Methodik werden in den nachfolgenden Abschnitten 5.4 bis 5.7 im Detail besprochen.



**Abb. 5-1:** Hauptphasen der Entwurfsmethodik für GIS-Visualisierungen

Ausgangspunkt der Methodik ist die Kompositionsalgebra von Mackinlay (siehe Abschnitt 3.2.1), die bereits in den verschiedensten intelligenten Visualisierungssystemen erfolgreich eingesetzt wurde [Mack86a, RoMa90, Casn91, SeIg94]. Mackinlay entwickelte seinen Ansatz ursprünglich für das wissensbasierte System APT zur automatisierten Präsentation relationaler Datensätze in Form von Diagrammen und Grafiken. Im Vergleich zu den Anforderungen an die intelligente GIS-Visualisierungsumgebung, waren die Ansprüche an APT bescheidener. Insbesondere folgende Anforderungen werden durch die Entwurfsmethodik von APT nicht erfüllt:

- Auswahl von Visualisierungen unter Beachtung der Effektivität bezüglich Zielsetzung, Ausgabemedium, Benutzer, etc.
- Vorschlag mehrerer alternativer Entwürfe, die bezüglich ihrer Effektivität geordnet sind
- globale Optimierung von Visualisierungsparametern und Klassenaufteilungen

Abb. 5-1 zeigt die Hauptphasen einer erweiterten Entwurfsmethodik, die auch den zusätzlichen Anforderungen der GIS-Visualisierung gerecht wird. Sie unterscheidet sich von Mackinlays Ansatz zunächst durch die gleichzeitige Verarbeitung mehrfacher Entwürfe (in der Abbildung durch die parallelen Pfeile angedeutet), durch eine zwischengeschaltete Phase der Parameteroptimierung und durch eine explizite, abschließende Phase der Effektivitätsabschätzung. Die Methodik umfaßt insgesamt vier Hauptphasen:

**1. Primitiver Visualisierungsentwurf:** Der Datensatz wird zunächst mit den Ausdrucksregeln aller verfügbaren Visualisierungstechniken verglichen. Jedesmal, wenn eine Teilrelation des Datensatzes die Ausdruckskriterien einer Technik erfüllt, wird eine Visualisierung erzeugt. Die so erzeugten Visualisierungen heißen *primitiv*, da sie jeweils nur genau eine Visualisierungstechnik verwenden. Sie heißen zusätzlich *partiell*, wenn sie nur einen Teil des Datensatzes ausdrücken, bzw. *vollständig*, wenn sie den gesamten Datensatz ausdrücken.

**2. Visualisierungskomposition:** In dieser Phase werden die vorher erzeugten primitiven Visualisierungen schrittweise zusammengesetzt. Dazu stehen eine Reihe von Kompositionsoptionen zur Verfügung, deren Anwendbarkeit auf Visualisierungen durch Kompositionsregeln bestimmt wird. Durch diese Regeln wird unter anderem sichergestellt, daß nur solche Visualisierungen zusammengesetzt werden, die unterschiedliche Teile des Datensatzes ausdrücken. Sind die durch einen Kompositionsschritt entstandenen *zusammengesetzten* Visualisierungen noch nicht vollständig, werden sie in den Kompositionsprozeß zurückgeführt. Dies wird so lange wiederholt, bis keine weiteren Kompositionen durchgeführt werden können und folglich alle vollständigen Visualisierungen des Datensatzes gefunden sind.

**3. Visualisierungsparameter-Optimierung:** In der dritten Phase wird jede vollständige Visualisierung traversiert und die offenen Parameter aller enthaltenen primitiven Visualisierungen bestimmt. Zu den Parametern von Visualisierungstechniken zählen vor allem die visuelle Ausgestaltung mit Signaturen und Farben sowie die Datenklassifizierung bei quantitativen Techniken. Unter Verwendung entsprechender Regeln werden die Parameter für jede vollständige Visualisierung optimiert und zugewiesen.

**4. Effektivitätsabschätzung:** Als abschließende Phase folgt eine Abschätzung der Effektivität aller vollständigen Visualisierungen. Dazu werden sie erneut traversiert und die Effektivität der enthaltenen primitiven Visualisierungen sowie die Effektivität der durchgeführten Kompositionen bestimmt. Es stehen Regeln zur Verfügung, die die Effektivität von Visualisierungen bezüglich der verschiedenen Faktoren wie Visualisierungszielsetzung, Medium, Zielgruppe, etc. bewerten. Mit einem Auswertungsschema wird aus diesen partiellen Effektivitätswerten eine Abschätzung über die Gesamteffektivität jeder vollständigen Visualisierung gewonnen. Danach lassen sich die Visualisierungen bezüglich ihrer Effektivität in eine eindeutige Reihenfolge bringen.

Betrachtet man die Reihenfolge der vier Phasen dieser Methodik, stellt sich folgende Frage: warum erfolgt die Parameterzuweisung und Effektivitätsabschätzung primitiver Visualisierungen erst im Anschluß an die Kompositionsphase? Eine optimale Zuweisung der Parameter einer Teilvisualisierung kann erst dann erfolgen, wenn bekannt ist, mit welchen Kompositionsoptionen sie in die vollständige Visualisierung aufgenommen

wird. Werden z.B. zwei Choroplethen überlagert, muß sichergestellt sein, daß die verwendeten Farbsignaturen auch in der Überlagerung noch erkennbar und eindeutig zuzuordnen sind. Die Effektivitätsabschätzung schließlich kann erst nach der Zuweisung der Visualisierungsparameter erfolgen, da die Wahl der Parameter wie Farben, Signaturen und Klassenaufteilungen einen Einfluß auf die Effektivität hat.

## 5.4 Primitiver Visualisierungsentwurf

Dieser Abschnitt diskutiert die erste Phase der Entwurfsmethodik, in der alle primitiven Teilvisualisierungen eines raumbezogenen Datensatzes erzeugt werden. Vorher werden noch die raumbezogenen Datensätze formal definiert und einige relevante Begriffe aus der Algebra definiert.

### 5.4.1 Raumbezogene Datensätze

#### Relationen

Wir betrachten Relationen der Form

$$R : (U_1 \times \dots \times U_n) \rightarrow (A_1 \times \dots \times A_m); \quad n, m \geq 1;$$

also Abbildungen von Mengen  $U = (U_1 \times \dots \times U_n)$  in Mengen  $A = (A_1 \times \dots \times A_m)$ .  $U$  heißt *unabhängiger Bereich*,  $A$  *abhängiger Bereich* von  $R$ . Nach dem Grad  $m$  des abhängigen Bereichs wird  $R$  *univariat* (falls  $m = 1$ ), *bivariat* (falls  $m = 2$ ) oder *multivariat* (falls  $m > 2$ ) genannt.

Für die Relation  $R$  wird im allgemeinen nicht vorausgesetzt, daß sie auf ganz  $U$  definiert ist, der Definitionsbereich kann also kleiner als der unabhängige Bereich sein. Gilt jedoch zusätzlich  $\mathbf{D}(R) = U$ , heißt  $R$  *vollständig*.

Anstelle der ausführlichen Schreibweise wird im folgenden immer die kürzere Notation

$$R : U_1, \dots, U_n \rightarrow A_1, \dots, A_m$$

verwendet.

#### Variablen

Der unabhängige Bereich  $U$  einer Relation  $R$  ist ein Produkt aus Wertemengen  $U_i$  ( $1 \leq i \leq n$ ). Da es sich bei  $R$  um eine Abbildung handelt, sind in jedem Tupel  $(u_1, \dots, u_n, a_1, \dots, a_m) \in R$  die Werte  $a_j$  vollständig durch die  $u_i$  bestimmt. Die Wertemengen  $U_i$  entsprechen dabei jeweils einem Freiheitsgrad von  $R$  und werden als die unabhängigen Variablen von  $R$  bezeichnet. Analog werden die  $A_j$  als die abhängigen Variablen von  $R$  bezeichnet.

Die Bezeichnung Variable wird in diesem Kontext also nicht mit der Bedeutung „Platzhalter“, sondern mit der Bedeutung „Variationsbereich“ verwendet. Die Repräsentation

einer Variable als Software-Objekt wird außer dem Wertebereich noch andere Komponenten umfassen, beispielsweise den Variablennamen und den Ordnungstyp. Variablen werden nach ihrem Ordnungstyp entweder als nominal, ordinal, quantitativ oder geographisch bezeichnet (siehe hierzu die Definitionen in Abschnitt 4.1.2).

### Raumbezogene Datensätze

Ein Datensatz  $D = \{R_1, \dots, R_k\}$  kann aus einer oder mehreren Relationen bestehen. Eine Relation heißt *raumbezogen*, wenn mindestens eine ihrer Variablen geographisch ist. Analog heißen Datensätze *raumbezogen*, wenn mindestens eine der enthaltenen Relationen *raumbezogen* ist.

Raumbezogene Datensätze werden gewöhnlich eingesetzt, um Charakteristika von Objekten der realen Welt und Zustände der realen Welt zu beschreiben. Aus diesem Grund treten sie sehr häufig in Geo-Informationssystemen nach Abfrage- und Analyseoperationen auf. Datensätze mit zweidimensionalem Raumbezug bilden den Schwerpunkt der hier vorgestellten Visualisierungsmethodik.

### Beispiel 5-1

Wir betrachten als Beispiel den Datensatz  $\{Klima\}$ , dessen Charakteristika schon in Beispiel 4-1 formal beschrieben wurden. Er besteht aus einer einzigen *raumbezogenen* Relation

$$Klima : Bezirk, Monat \rightarrow Niederschlag, Temperatur$$

mit den unabhängigen Variablen *Bezirk* und *Monat*, sowie den abhängigen Variablen *Niederschlag* und *Temperatur*. *Bezirk* ist geographisch, *Monat* ordinal und die beiden abhängigen Variablen sind jeweils quantitativ.

<i>Klima:</i>			
<i>Bezirk</i>	<i>Monat</i>	<i>Niederschlag</i>	<i>Temperatur</i>
A	Januar	34	6
A	Februar	22	9
...	...	...	...
Z	Dezember	113	20

## 5.4.2 Dekomposition von Relationen

Die Mehrzahl der primitiven GIS-Visualisierungstechniken können lediglich Relationen mit einer oder zwei unabhängigen und einer oder zwei abhängigen Variablen ausdrücken. Choroplethen und Kreissignaturen sind beispielsweise für eine unabhängige, geographische und eine abhängige, quantitative Variable definiert. Eine Ausnahme von dieser Beschränkung bilden nur die Gesichterdarstellungen und vergleichbare Techniken, die multivariate Relationen direkt ausdrücken. Ein automatisches Entwurfssystem muß deshalb komplexe Relationen in eine Anzahl von *Teilrelationen* zerlegen, die Ausdruckskriterien

primitiver Techniken erfüllen. Primitive Visualisierungen dieser Teilrelationen können anschließend zu Visualisierungen der vollständigen Relation zusammengesetzt werden.

Um ausdrucksfähige Visualisierungen mit diesem Ansatz zu erhalten, muß die Dekomposition der Relation und auch die Komposition der Visualisierung *informationserhaltend* sein. Eine Dekomposition einer Relation heißt informationserhaltend, wenn die Summe der in den Teilrelationen enthaltenen Information gleich der in der ursprünglichen Relation enthaltenen Information ist. Nur wenn alle Dekompositions- und Kompositionsschritte der Methodik informationserhaltend sind, gibt die Visualisierung genau die Information des Datensatzes wieder und ist per Definition ausdrucksfähig.

### Funktionale Dekomposition

Eine einfache Technik zur Dekomposition von Relationen ist die funktionale Dekomposition: sie spaltet die Menge der abhängigen Variablen einer Relation auf. Hat die Relation die Funktionseigenschaft, ist die Dekomposition gleichzeitig auch informationserhaltend. So läßt sich zum Beispiel die Relation

$$\text{Klima} : \text{Bezirk}, \text{Monat} \rightarrow \text{Niederschlag}, \text{Temperatur}$$

in zwei Relationen der Form

$$\text{Klima}_1 : \text{Bezirk}, \text{Monat} \rightarrow \text{Niederschlag}$$

$$\text{Klima}_2 : \text{Bezirk}, \text{Monat} \rightarrow \text{Temperatur}$$

aufspalten, ohne daß Information verloren geht. Diese Eigenschaft ist ein Korollar der Vereinigungs- und Dekompositionsregeln, die ihrerseits aus den Armstrong'schen Datenbankaxiomen folgen [Mack86a, KoSi86].

### Definitionsbereich-Dekomposition

Eine Dekompositionstechnik für die unabhängigen Variablen einer Relation ist die Definitionsbereich-Dekomposition. Dazu wird eine unabhängige Variable  $U$  der Relation ausgewählt und die Relation durch  $|U|$  Teilrelationen ersetzt. Jede dieser neuen Relationen wird gebildet, indem man aus der ursprünglichen Relation alle Tupel auswählt, die einen bestimmten, festen Wert für  $U$  haben und anschließend  $U$  aus diesen Tupeln entfernt. Die Relation *Klima* kann auf diese Art in zwölf Teilrelationen der Form

$$\text{Bezirk}_{\text{Jan}} : \text{Land} \rightarrow \text{Niederschlag}, \text{Temperatur}$$

$$\text{Bezirk}_{\text{Feb}} : \text{Land} \rightarrow \text{Niederschlag}, \text{Temperatur}$$

...

$$\text{Bezirk}_{\text{Dez}} : \text{Land} \rightarrow \text{Niederschlag}, \text{Temperatur}$$

zerlegt werden. Die Teilrelationen enthalten jeweils die Niederschlags- und Temperaturwerte aller Länder für einen bestimmten Monat.

Die Definitionsbereichs-Dekomposition alleine ist noch nicht informationserhaltend [Mack86a]. Die zwölf Teilrelationen, die durch Dekomposition von *Klima* gebildet wur-

den, enthalten nicht mehr die Information, für welchen Monat sie die Werte wiedergeben. Um alle Information der ursprünglichen Relation zu erhalten, müssen die Teilrelationen jeweils noch zusätzlich mit der Information ausgestattet werden, welche Variable der ursprünglichen Relation sie festsetzen und auf welchen Wert sie sie setzen. Die Relation  $Bezirk_{Jan}$  muß z.B. zusätzlich mit der Information  $Monat = \text{Januar}$  ausgestattet werden.

Mit funktionaler und Definitionsbereich-Dekomposition läßt sich die Anzahl unabhängiger und abhängiger Variablen multivariater Relationen verringern. Zu diesen beiden Dekompositionen können außerdem komplementäre Kompositionsoperationen (siehe Abschnitt 5.5) definiert werden, die für die Visualisierung raumbezogener Datensätze geeignet sind. Mackinlay hat in seinem Ansatz noch eine dritte Dekomposition, die Bildbereich-Dekomposition aufgenommen. Sie wird dann nötig, wenn auch multivariate nicht-Funktionen zerlegt werden sollen, erfordert jedoch die Aufnahme einer sehr allgemeinen Kompositionsoperation, die in GIS-Visualisierungskomponenten nicht zur Verfügung steht. Die Bildbereich-Dekomposition wurde deswegen nicht in die Methodik für den GIS-Visualisierungsentwurf aufgenommen. Auch andere automatisierte Präsentationssysteme, die auf APT aufbauen, haben aus ähnlichen Gründen auf die Bildbereich-Dekomposition verzichtet [BeFe93].

### 5.4.3 Matching und Primitiver Entwurf

Um effektive Visualisierungen einer Relation zu finden, müssen im allgemeinen alle Dekompositionen der Relation betrachtet werden. Sollen alle Teilrelationen einer Relation gebildet und auf passende primitive Visualisierungstechniken untersucht werden, ist ein enormer Rechenaufwand nötig. Eine Relation mit  $n$  unabhängigen und  $m$  abhängigen Variablen kann in

$$2^m \sum_{k=0}^n \binom{n}{k} c^k$$

verschiedene Teilrelationen zerlegt werden, unter der vereinfachenden Annahme, daß alle unabhängigen Variablen die gleiche Anzahl  $c$  von Werten haben. Für eine Relation mit 3 unabhängigen Variablen von je 20 Werten und mit 3 abhängigen Variablen ergeben sich schon mehr als 74 000 verschiedene Teilrelationen.

Der Suchraum für primitive Visualisierungen kann jedoch signifikant verringert werden. Normalerweise erwarten wir für alle Teilrelationen, die aus einer Definitionsbereich-Dekomposition stammen, daß sie auf die gleiche Art visualisiert werden. So wäre es normalerweise störend, die Teilrelationen  $Bezirk_{Jan}, \dots, Bezirk_{Dez}$  mit verschiedenartigen Visualisierungen darzustellen, da die daraus zusammengesetzte Visualisierung von *Klima* ungewöhnlich wäre und den Vergleich von Werten für verschiedene Monate erschweren würde. Setzt man nun die Verwendung gleichartiger Visualisierungen für alle derartigen Teilrelationen voraus, genügt es ausdrucksfähige Visualisierungen für einen *Repräsentanten* der Teilrelationen zu entwerfen.



Der Entwurf kann anschließend für alle Teilrelationen, die aus derselben Definitionsbereich-Dekomposition stammen, verwendet werden. Die Verwendung von Repräsentanten verringert die Anzahl der zu untersuchenden Teilrelationen auf  $2^{m+n}$ , da der Faktor  $c^k$  im obigen Ausdruck wegfällt. Für eine Relation mit 3 unabhängigen und 3 abhängigen Variablen entstehen z.B. lediglich 64 Repräsentanten.

Anstatt die Relation explizit in Teilrelationen und Repräsentanten zu zerlegen, kann die Dekomposition auch implizit beim Berechnen der *Partial-Matches* (partiellen Übereinstimmungen) zwischen der Relation und den Ausdruckskriterien der primitiven Techniken durchgeführt werden. In einem Partial-Match werden alle Ausdruckskriterien einer Visualisierungstechnik  $V$  von Variablen der Relation  $R$  erfüllt, aber einige der Variablen von  $R$  sind möglicherweise nicht in  $V$  dargestellt. Partial-Matches werden als Prädikate

$$\text{Vis}(V, U, A)$$

notiert. Hierbei geben  $U$  und  $A$  die Menge derjenigen unabhängigen bzw. abhängigen Variablen von  $R$  wieder, die in  $V$  dargestellt sind. Jeder Partial-Match entspricht einer primitiven Visualisierung eines Teilrelation-Repräsentanten von  $R$ . Er impliziert eine funktionale Dekomposition von  $R$ , die alle in  $A$  enthaltenen abhängigen Variablen von denjenigen Variablen abspaltet, die nicht in  $A$  enthalten sind. Er impliziert außerdem je eine Definitionsbereich-Dekomposition für jede unabhängige Variable  $U$ , die nicht in  $U$  enthalten ist. Das Verfahren der Partial-Matches mit impliziter Dekomposition erspart das Erzeugen von Teilrelationen und Repräsentanten, die von keiner der vorhandenen primitiven Visualisierungstechniken ausgedrückt werden kann.

#### 5.4.4 Regeln zur Ausdrucksfähigkeit

##### Regelnotation

Die Notation für Regeln in Kapitel 5 stellt eine starke Abstraktion von der Syntax üblicher regelbasierter Expertensysteme dar. Durch diese Abstraktion wird eine Einschränkung auf eine bestimmte Expertensystemshell sowie die ausführliche Definition der entsprechenden Syntax vermieden. Die Notation enthält jedoch die wichtigsten Konstrukte aus Expertensystemen und es bleibt nachvollziehbar, wie die Regeln in einer bestimmten Shell implementiert werden können.

Regeln haben einen **IF**- und einen **THEN**-Teil. Der **IF**-Teil, häufig auch LHS (left hand side) genannt, besteht aus einem Boole'schen Ausdruck. Der **THEN**-Teil enthält Anweisungen der Form **ASSERT Prädikat** und **RETRACT Prädikat**, die ein Prädikat in den Faktenspeicher des Systems aufnehmen bzw. es daraus entfernen. Die Auswertung eines Prädikates (z.B. in der LHS) ergibt genau dann den Wert „wahr“, wenn das Prädikat im Faktenspeicher enthalten ist.

##### Beispiel 5-2

Für die Relation *Klima* soll der Matchingprozeß durchgeführt und die primitiven Visualisierungen erzeugt werden. Dazu betrachten wir beispielhaft einen eingeschränkten Satz von vier Visualisierungstechniken: Choroplethenkarten, Kreissignaturkarten, bivariate

Choroplethenkarten und Tortendiagrammkarten. Die Ausdruckskriterien dieser Techniken wurden bereits in den Abschnitten 2.4.2, 2.4.3 und 2.4.6 vorgestellt und werden durch die folgenden vier Regeln formalisiert:

**Regel 5-1** (Ausdruckskriterien der Choroplethenkarte)

**IF** die Relation ist eine Funktion  $U \rightarrow A$

- $\wedge$   $U$  ist eine geographische Variable aus zwei-dimensionalen Geo-Objekten oder Geo-Referenzen
- $\wedge$   $A$  ist eine quantitative Variable, die bezüglich der Fläche normalisiert oder unabhängig von der Fläche ist

**THEN ASSERT** Vis(Choroplethen( $U \rightarrow A$ ),  $\{U\}$ ,  $\{A\}$ )

**Regel 5-2** (Ausdruckskriterien der Kreissignaturenkarte)

**IF** die Relation ist eine Funktion  $U \rightarrow A$

- $\wedge$   $U$  ist eine geographische Variable aus zwei-dimensionalen Geo-Objekten oder Geo-Referenzen
- $\wedge$   $A$  ist eine quantitative Variable

**THEN ASSERT** Vis(Kreissignaturen ( $U \rightarrow A$ ),  $\{U\}$ ,  $\{A\}$ )

**Regel 5-3** (Ausdruckskriterien der bivariaten Choroplethenkarte)

**IF** die Relation ist eine Funktion  $U \rightarrow A_1, A_2$

- $\wedge$   $U$  ist eine geographische Variable aus zwei-dimensionalen Geo-Objekten oder Geo-Referenzen
- $\wedge$   $A_1$  und  $A_2$  sind quantitative Variablen, die bezüglich der Fläche normalisiert oder unabhängig von der Fläche sind

**THEN ASSERT** Vis(BiChoroplethen( $U \rightarrow A$ ),  $\{U\}$ ,  $\{A_1, A_2\}$ )

**Regel 5-4** (Ausdruckskriterien der Tortendiagrammkarte)

**IF** die Relation ist eine Funktion  $U_1, U_2 \rightarrow A$

- $\wedge$   $U_1$  ist eine geographische Variable aus 0D- oder 2D- Geo-Objekten oder Geo-Referenzen
- $\wedge$   $U_2$  ist eine nominale oder ordinale Variable
- $\wedge$   $A$  ist eine quantitative Variable
- $\wedge$   $U_2$  ist Partitionsindex zu  $A$  (siehe Definition auf Seite 27)

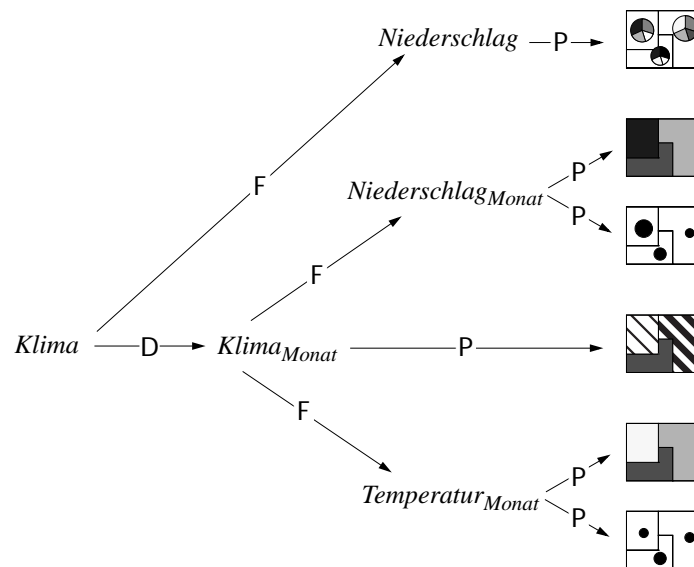
**THEN ASSERT** Vis(Tortendiagramme ( $U_1, U_2 \rightarrow A$ ),  $\{U_1, U_2\}$ ,  $\{A\}$ )

Zu den ersten beiden Regeln (Choroplethen, Kreissignaturen) lassen sich mehrere Teilrelationen von *Klima* finden, die die Ausdruckskriterien erfüllen. *Bezirk*  $\rightarrow$  *Niederschlag*, *Temperatur* wird durch die bivariate Choroplethenkarte ausgedrückt. *Monat* ist Partitionsindex zu *Niederschlag*, jedoch nicht zu *Temperatur* (die Durchschnittstemperaturen lassen sich im Gegensatz zu den Niederschlagsmengen nicht über die Monate summieren). Demnach wird *Bezirk, Monat*  $\rightarrow$  *Niederschlag* durch die Tortendiagrammkarte ausge-

drückt. Insgesamt werden folgende primitive Visualisierungen der Relation *Klima* gefunden:

$\text{Vis}(\text{Choroplethen } (\text{Bezirk} \rightarrow \text{Niederschlag}), \{\text{Bezirk}\}, \{\text{Niederschlag}\})$   
 $\text{Vis}(\text{Choroplethen } (\text{Bezirk} \rightarrow \text{Temperatur}), \{\text{Bezirk}\}, \{\text{Temperatur}\})$   
 $\text{Vis}(\text{Kreissignaturen } (\text{Bezirk} \rightarrow \text{Niederschlag}), \{\text{Bezirk}\}, \{\text{Niederschlag}\})$   
 $\text{Vis}(\text{Kreissignaturen } (\text{Bezirk} \rightarrow \text{Temperatur}), \{\text{Bezirk}\}, \{\text{Temperatur}\})$   
 $\text{Vis}(\text{BiChoroplethen } (\text{Bezirk} \rightarrow \text{Niederschlag}, \text{Temperatur}), \{\text{Bezirk}\}, \{\text{Niederschlag}, \text{Temperatur}\})$   
 $\text{Vis}(\text{Tortendiagramme } (\text{Bezirk}, \text{Monat} \rightarrow \text{Temperatur}), \{\text{Bezirk}, \text{Monat}\}, \{\text{Temperatur}\})$

Diese Partial-Matches implizieren zwei Dekompositionen von *Klima*: eine Definitionsbereich-Dekomposition bezüglich der Variable *Monat* und eine funktionale Dekomposition, die die Variablen *Niederschlag* und *Temperatur* trennt.  $\text{Bezirk} \rightarrow \text{Niederschlag}$  ist beispielsweise ein Repräsentant der zwölf Teilrelationen, die Niederschlagswerte für jeweils einen Monat enthalten (siehe Abb. 5-2).



**Abb. 5-2:** Erste Phase des Visualisierungsentwurfs für die Relation *Klima* : funktionale (F) bzw. Definitionsbereich-Dekomposition (D) und primitive Visualisierung (P).

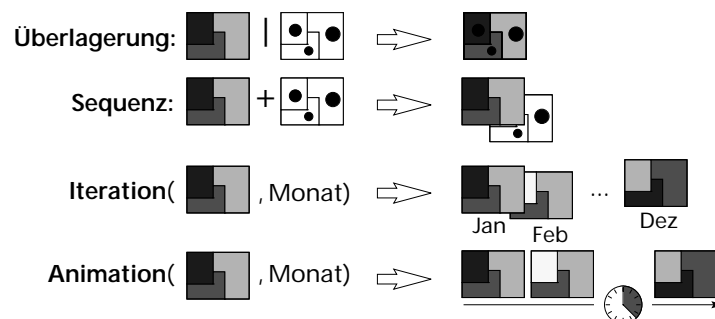
## 5.5 Komposition von Visualisierungen

Sollen partielle Visualisierungen durch eine Komposition zusammengesetzt werden, müssen eine Reihe von Bedingungen erfüllt sein. Einerseits muß sichergestellt sein, daß die entstehende komplexe Visualisierung widerspruchsfrei, übersichtlich und effektiv ist. Andererseits müssen aber auch die durch die Visualisierungskomposition implizit zusam-

mengefügten Teilrelationen zusammenpassen. Es dürfen nur Visualisierungen disjunkter Teilrelationen zusammengesetzt werden, da sonst die komplexe Visualisierung Information mehrfach ausdrücken würde. Die Operationen zur Visualisierungskomposition lassen sich so wählen, daß sie jeweils zu einer der Dekompositionsooperatoren komplementär sind. Sie fügen dann Visualisierungen von Teilrelationen zusammen, die aus einer vorhergehenden Dekomposition stammen und folglich disjunkt sind.

Zur Definition der Kompositionsoperationen muß weiterhin beachtet werden, daß sie einen ausreichend großen Entwurfsraum für GIS-Visualisierungen aufspannen. Um unterschiedliche Zielsetzungen zu erfüllen, müssen z.B. einerseits Operationen definiert werden, die die Information möglichst kompakt darstellen und einen globalen Vergleich zwischen mehreren Variablen ermöglichen. Es sind aber auch solche Operationen nötig, die den Blick auf das Detail dadurch freihalten, daß sie jeder Variablen ihren eigenen Darstellungsbereich gewähren.

Ausgehend von diesen Betrachtungen definiert Abschnitt 5.5 zunächst vier Operatoren zur Visualisierungskomposition: Überlagerung, Sequenz, Iteration und Animation (siehe Abb. 5-3). Überlagerung und Sequenz sind komplementär zur funktionalen Dekomposition, Iteration und Animation zur Definitionsbereich-Dekomposition. Diese vier Operatoren spannen zusammen den für die GIS-Visualisierung fundamentalen Entwurfsraum der (2D-) thematischen Karten und kartenverwandten Darstellungen auf. Anhand dieser Operatoren soll in Abschnitt 5.5 die Visualisierungskomposition im Detail dargestellt werden.



**Abb. 5-3:** Kompositionsoperationen für thematische Karten

### 5.5.1 Überlagerung

Die Überlagerung ist eine fundamentale kartographische Arbeitstechnik für die Analyse bivariater und multivariater raumbezogener Datensätze. Begriff und Konzept der Überlagerung stammen aus der Zeit der manuellen Kartographie. Überlagerung im GIS impliziert jedoch neben der rein graphischen Superimposition oft auch eine algorithmische Verschränkung der Datensätze, z.B. mittels Polygonverschneidung.

Eine Überlagerung zweier Karten setzt eine Reihe von Bedingungen voraus, die in der entsprechenden Kompositionsregel überprüft werden müssen. Zunächst können Karten

nur dann überlagert werden, wenn sich ihre Maßstabsbereiche  $Scale(V)$  überlappen. Für die graphische Superimposition sind die Signatordimensionen der Karten relevant: im allgemeinen können punktförmige und linienförmige Signaturen einer flächenförmigen Signatur überlagert werden. Im umgekehrten Fall würde die flächenförmige Signatur die anderen Signaturen jedoch überdecken. Zwei flächenförmige Signaturen wiederum können nur überlagert werden, wenn mindestens eine davon transparent ausgestaltet ist, z.B. durch ein Linienraster.

Als Kompositionsoperation fügt die Überlagerung zwei aus einer funktionalen Dekomposition entstandene Teilrelationen zusammen. Dies ist als weitere Bedingung derart zu formulieren, daß die Teilrelationen den gleichen unabhängigen Bereich und disjunkte abhängige Bereiche haben müssen. Insgesamt ergibt sich folgende Regel für die Überlagerung  $V_1|V_2$  der partiellen Visualisierung  $V_2$  über die partielle Visualisierung  $V_1$ :

**Regel 5-5** (Ausdrucksfähigkeit der Überlagerung)

**IF**  $Vis(V_1, U, A_1)$   
 $\wedge Vis(V_2, U, A_2)$   
 $\wedge A_1 \cap A_2 = \emptyset$   
 $\wedge V_1$  und  $V_2$  sind Karten  
 $\wedge Scale(V_1) \cap Scale(V_2) \neq \emptyset$   
 $\wedge V_2$  kann  $V_1$  graphisch überlagert werden  
**THEN ASSERT**  $Vis(V_1|V_2, U, A_1 \cup A_2)$

Einige der graphischen Bedingungen in Regel 5-5 können unter bestimmten Voraussetzungen noch abgeschwächt werden. Verfügt die GIS-Visualisierungssoftware über ein Modul zur Generalisierung, sind dynamische Erweiterungen des Maßstabsbereichs und weitergehende Signaturüberlagerungen möglich. Eine solche Funktionalität kann aber erst in einer der nächsten Generationen der GIS erwartet werden [FrTi94].

## 5.5.2 Sequenz

Die Sequenz fügt, analog zur Überlagerung, Visualisierungen von Teilrelationen aus einer funktionalen Dekomposition zusammen. Sie überlagert die partiellen Visualisierungen jedoch nicht in einem gemeinsamen Fenster, sondern stellt jede in einem separaten Fenster dar. Sie werden also nicht graphisch, sondern lediglich konzeptionell zu einer komplexeren Visualisierung zusammengefügt.

Die Regel zur Ausdrucksfähigkeit der Sequenz ist entsprechend einfach. Es gibt keine Einschränkungen bezüglich Maßstabsbereichen oder Signatordimensionen:

**Regel 5-6** (Ausdrucksfähigkeit der Sequenz)

**IF**  $\text{Vis}(V_1, U, A_1)$   
 $\wedge \text{Vis}(V_2, U, A_2)$   
 $\wedge A_1 \cap A_2 = \emptyset$   
**THEN ASSERT**  $\text{Vis}(V_1 + V_2, U, A_1 \cup A_2)$

### Beispiel 5-3

Im letzten Beispiel wurden je zwei primitive Visualisierungen von *Bezirk*  $\rightarrow$  *Niederschlag* und von *Bezirk*  $\rightarrow$  *Temperatur* erzeugt. Beide Visualisierungen von *Bezirk*  $\rightarrow$  *Niederschlag* sind mit beiden Visualisierungen von *Bezirk*  $\rightarrow$  *Temperatur* sowohl durch Überlagerung als auch durch Sequenz zusammensetzbar. Es entsteht jeweils eine Visualisierung von *Bezirk*  $\rightarrow$  *Niederschlag, Temperatur*, so z.B.:

$$\text{Vis}(\text{Choroplethen}(\text{Bezirk} \rightarrow \text{Temperatur}) | \text{Kreissignaturen}(\text{Bezirk} \rightarrow \text{Niederschlag}), \{\text{Bezirk}\}, \{\text{Niederschlag, Temperatur}\})$$

Mit der Tortendiagrammkarte kann jedoch keine der anderen primitiven Visualisierungen zusammengesetzt werden, da die Bedingung an den unabhängigen Bereich nicht erfüllt wird.

### 5.5.3 Iteration

Analog zur Sequenz stellt die Iteration jede partielle Visualisierung in einem separaten Fenster dar. Die Iteration setzt jedoch Teilrelationen aus einer Definitionsbereich-Dekomposition zusammen. Diese Teilrelationen werden, wie in 5.4.2 beschrieben, durch einen Repräsentanten verkörpert. Visualisierungen von Repräsentanten lassen sich daran erkennen, daß sie mindestens eine unabhängige Variable der Relation nicht darstellen. So stellt die Überlagerung von Kreissignaturen auf Choroplethen aus Beispiel 5-3 die unabhängige Variable *Monat* nicht dar. Sie ist also Visualisierung eines Repräsentanten und kann iteriert werden.

Die Iteration wiederholt die Visualisierungsform des Repräsentanten für alle durch ihn verkörperten Teilrelationen. Dabei werden alle Werte der betroffenen unabhängigen Variablen *U* durchlaufen und je ein Visualisierungsfenster erzeugt. Wir sprechen deswegen auch von der Iteration einer Visualisierung *V* für eine unabhängige Variable *U*. Im Beispiel werden alle Werte der Variablen *Monat* durchlaufen und für *Monat* = Januar, *Monat* = Februar, ..., *Monat* = Dezember je ein Fenster erzeugt.

Die Kompositionsregel für die Iteration lautet:

**Regel 5-7** (Ausdrucksfähigkeit der Iteration)

**IF**  $\text{Vis}(V, U, A)$   
 $\wedge U$  ist eine unabhängige Variable der Relation  
 $\wedge U \notin U$   
 $\wedge \#(V) \cdot |U| \leq \#_{\text{Max}}$   
**THEN ASSERT**  $\text{Vis}(\text{Iteration}(V), U \cup \{U\}, A)$

Durch mehrfache Iteration kann die Anzahl der Fenster einer Visualisierung stark ansteigen. Bei der Iteration einer Visualisierung  $V$  für eine Variable  $U$  wird die Anzahl  $\#(V)$  der Fenster um den Faktor  $|U|$  vergrößert. Die Regel enthält deswegen eine obere Schranke  $\#_{\text{Max}}$  für die Anzahl der Fenster, die auf diesem Weg erzeugt werden kann. Diese Schranke kann jedoch entfallen, wenn das Visualisierungstool einen Browser enthält, mit dem Benutzer auf einzelne Fenster direkt zugreifen können.

**5.5.4 Animation**

Die Animation setzt ebenfalls Teilrelationen aus einer Definitionsbereich-Dekomposition zusammen. Im Gegensatz zur Iteration werden jedoch keine separaten Fenster erzeugt, sondern die Teilvisualisierungen bilden die Einzelbilder einer Animationssequenz. Jedes Bild der Animation korrespondiert mit genau einem Wert aus dem Wertebereich von  $U$  und alle Bilder verwenden die gleiche Visualisierungsform.

Für die Zulässigkeit der Animation gelten strengere Bedingungen. Visualisierungen können nur für ordinale oder quantitative Variablen animiert werden, da eine eindeutige, lineare Reihenfolge der Einzelbilder und damit auch der Variablenwerte gegeben sein muß. Aus den gleichen Gründen ist die Schachtelung von Animationen nicht erlaubt: wird eine primitive Visualisierung erst für eine Variable  $U_1$  und zusätzlich noch für eine Variable  $U_2$  animiert, ist die Abfolge der Einzelbilder nicht mehr linear und die Animationssequenz wirkt in der Regel sprunghaft. Auch die dargestellten abhängigen Variablen müssen bezüglich  $U$  einigermaßen stetig sein, damit sich ein glatter Übergang zwischen den Einzelbildern ergibt. Schließlich darf die resultierende Animationssequenz nicht zu kurz sein, es muß eine gewisse Mindestanzahl  $\#_{\text{Anim}}$  von Bildern erzeugt werden.

Die Kompositionsregel für die Animation lautet insgesamt:

**Regel 5-8** (Ausdrucksfähigkeit der Animation)

**IF**  $\text{Vis}(V, U, A)$   
 $\wedge V$  ist nicht animiert  
 $\wedge U$  ist eine unabhängige, ordinale oder quantitative Variable der Relation  
 $\wedge U \notin U$   
 $\wedge |U| \geq \#_{\text{Anim}}$   
 $\wedge \forall A \in A : A$  ist stetig auf  $U$   
 $\wedge$  das Ausgabemedium unterstützt Animation  
**THEN ASSERT**  $\text{Vis}(\text{Animation}(V), U \cup \{U\}, A)$

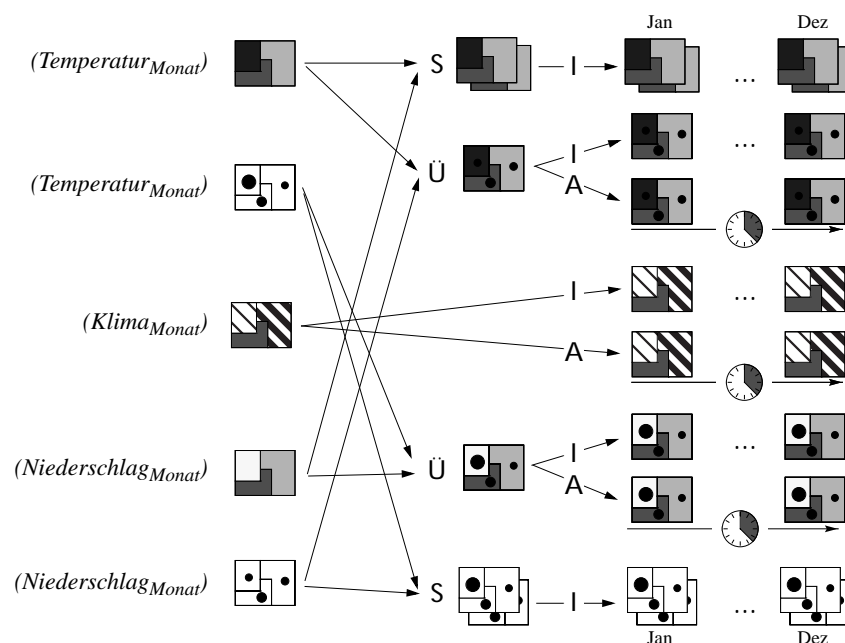
Diese Regel erlaubt auch die Animation von Visualisierungen mit mehreren Fenstern. Damit kann z.B. die zeitliche Entwicklung mehrerer Variablen parallel in nebeneinander-gestellten Fenstern beobachtet werden.

#### Beispiel 5-4

Vollständige Visualisierungen der Relation *Klima* können z.B. durch Iteration oder durch Animation der Überlagerung aus Beispiel 5-3 erzeugt werden:

Vis(Iteration(Choroplethen(*Bezirk* → *Temperatur*)|Kreissignaturen(*Bezirk* →  
*Niederschlag*), *Monat*), { *Bezirk*, *Monat* }, { *Niederschlag*, *Temperatur* } )  
 Vis(Animation(Choroplethen(*Bezirk* → *Temperatur*)|Kreissignaturen(*Bezirk* →  
*Niederschlag*), *Monat*), { *Bezirk*, *Monat* }, { *Niederschlag*, *Temperatur* } )

Abb. 5-4 zeigt alle Kompositionen (außer denen für die Tortendiagrammkarten), die beim Visualisierungsentwurf für die Relation *Klima* durchgeführt werden.



**Abb. 5-4:** Visualisierungskompositionsphase (Ausschnitt) für die Relation *Klima* : Überlagerung (Ü), Sequenz (S), Iteration (I) und Animation (A).

#### 5.5.5 Sukzessive Komposition

Die Methodik führt sukzessive alle Kompositionen partieller Visualisierungen des Datensatzes durch. Jedesmal, wenn in einem Schritt eine Kompositionsregel zutrifft, erzeugt sie mit der *ASSERT*-Anweisung eine neue Visualisierung. Diese neue Visualisierung steht dann für weitere Kompositionen zur Verfügung. Die Kompositionsphase terminiert, wenn in einem Schritt keine neuen Visualisierungen erzeugt werden konnten. Vollständige Vi-



sualisierungen einer Relation  $R : U_1, \dots, U_n \rightarrow A_1, \dots, A_m$  drücken alle unabhängigen und abhängigen Variablen von  $R$  aus:

$$\text{Vis}(V, \{U_1, \dots, U_n\}, \{A_1, \dots, A_m\})$$

## 5.6 Parameteroptimierung

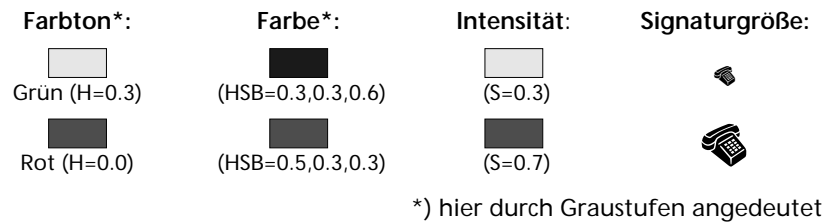
Die Visualisierungstechniken legen die in einer Visualisierung darzustellenden Signaturen nur zum Teil fest. Bei jeder Technik sind einige Parameter der graphischen Ausgestaltung, wie Signaturfarben und -größen variabel. Choroplethenkarten legen beispielsweise nur die Signaturdimension (flächenförmig) und -art (gestuft und ordinal) fest. Wie die gestufte Signatur ausgestaltet wird und wie die Daten klassifiziert werden, bleibt zunächst offen, beeinflusst jedoch im allgemeinen die Effektivität der Visualisierung. Diese Möglichkeiten der Ausgestaltung und Datenklassifikation werden als Parameter der Visualisierungstechnik bezeichnet.

Die Parameter der Visualisierung werden nach dem Ende der Kompositionsphase belegt, also nachdem alle vollständigen Visualisierungen des Datensatzes bestimmt sind. Die Parameterwahl kann erst nach der Komposition erfolgen, da die Struktur einer zusammengesetzten Visualisierung den Freiheitsgrad der Parameter einschränkt. So muß zum Beispiel bei einer Überlagerung von zwei Choroplethenkarten die obere transparent ausgestaltet sein. Die Parameterwahl muß aber *vor* der Effektivitätsabschätzung erfolgen, da die zugewiesenen Parameter die Effektivität der Visualisierung beeinflussen. Im Beispiel der überlagerten Choroplethen wird die obere weniger effektiv sein, da transparente flächenförmige Signaturen wie Punkte- oder Linienraster weniger effektiv als die opaken Intensitätssignaturen sind.

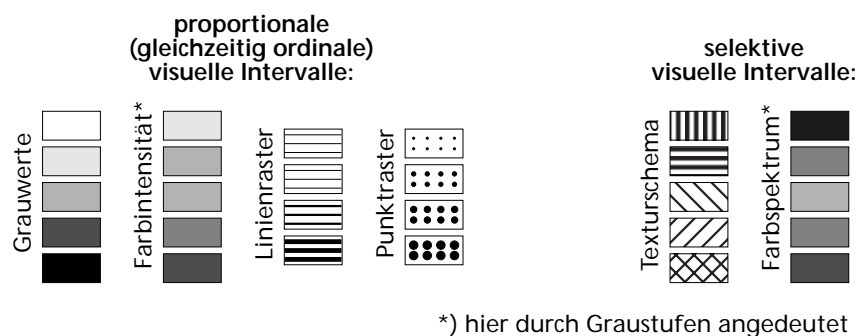
Für die Parameterwahl werden alle vollständigen Visualisierungen traversiert. Zu jeder enthaltenen primitiven Visualisierung werden die offenen Parameter mit ihren jeweiligen Freiheitsgraden bestimmt. Über die Belegung jedes Parameters entscheidet schließlich ein Ressourcenmanager, der den Einsatz der visuellen Variablen und Symbole optimiert und deren Konsistenz sicherstellt.

### 5.6.1 Parametertypen

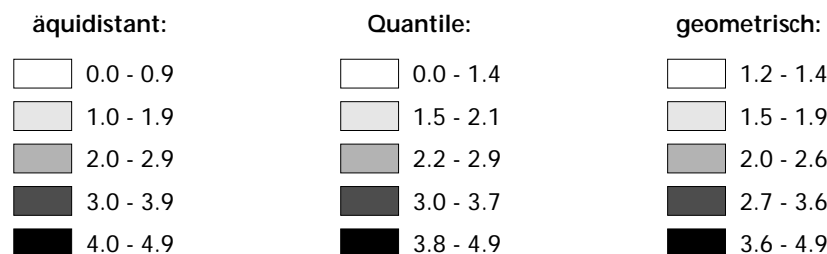
Visualisierungsparameter gestalten eine Signatur aus. Je nachdem, wie stark die Signatur schon durch die verwendete Visualisierungstechnik festgelegt ist, können die Parameter einzelne visuelle Konstanten, Kombinationen von Konstanten oder visuelle Intervalle sein. *Visuelle Konstanten* haben einen festen Wert aus dem Wertebereich einer oder mehrerer visueller Variablen (wie Farbton, Textur, Größe). Ein Beispiel einer solcher Konstante ist die Signaturfarbe als Parameter der Kreissignaturenkarte. Um eine Signaturfarbe zu definieren, werden drei visuelle Variablen, nämlich Farbton, Intensität und Helligkeit festgelegt. Abb. 5-5 zeigt je zwei Beispiele für die Belegung verschiedener Parameter vom Typ „visuelle Konstante“.



**Abb. 5-5:** Beispiele visueller Konstanten. Farbwerte sind im *HSB*-Schema (*Hue*, *Saturation*, *Brightness*) angegeben.



**Abb. 5-6:** Beispiele flächenförmiger visueller Intervalle



**Abb. 5-7:** Verschiedene Möglichkeiten der Datenklassifikation

*Visuelle Intervalle* haben Bedeutung für die Ausgestaltung gestufter und stetiger Signaturen. Ein visuelles Intervall belegt ein Intervall aus dem Wertebereich einer visuellen Variablen und häufig noch zusätzlich feste Werte aus anderen visuellen Variablen. Die gestufte, flächenförmige Signatur einer Choroplethenkarte kann z.B. mit einer Grauwert-rampe, einem Intervall der visuellen Variable Helligkeitswert, ausgestaltet werden. Abb. 5-6 zeigt einige Beispiele flächenförmiger visueller Intervalle. Es wird zwischen selektiven, ordinalen und proportionalen Intervallen unterschieden, je nach Wirkung der zugrundeliegenden visuellen Variablen (siehe hierzu Abb. 2-1).

Ein weiterer Parameter vieler quantitativer thematischer Karten ist die Klassenaufteilung. Klassifikationen unterteilen den Wertebereich einer quantitativen Variablen in eine feste, kleine Zahl von Klassen auf. Abb. 5-7 zeigt drei der häufigsten Klassifikationsverfahren.

Mit äquidistanten Klassen werden z.B. Höhendaten abgebildet, für statistische Daten sind sie jedoch ungeeignet. Zur Berechnung der Quantile wird ein Histogramm der Datenwerte so aufgeteilt, daß jede Klasse die gleiche Anzahl von Werten enthält. Ist eine Variable arithmetisch oder geometrisch verteilt, versucht man die Klassen entsprechend zu wählen.

### 5.6.2 Parameterregeln

Die Freiheitsgrade eines Parameters werden natürlich zunächst von der Visualisierungstechnik selbst bestimmt: Methoden für quantitative Variablen erlauben nur proportionale visuelle Intervalle, während für nominale Variablen ausschließlich selektive Intervalle sinnvoll sind. Jede Visualisierungstechnik erfordert deshalb eine Parameterregel, die aussagt, welche Parameter benötigt werden und wie sie jeweils beschaffen sein müssen.

Jeder einzelne Parameter einer Visualisierungstechnik  $V$  wird notiert als

$$\text{Param}(V, i, T(a_1, \dots, a_m)),$$

wobei  $i$  die Parameternummer wiedergibt.  $T$  ist der Parametertyp, also visuelle Konstante, visuelles Intervall oder Klassenaufteilung.  $a_1, \dots, a_m$  stellen die Attribute bzw. Freiheitsgrade des Parameters dar, so z.B. die Dimension, die geforderte visuelle Variable, etc.

#### Beispiel 5-5

Die Choroplethentechnik ist als quantitative Methode durch ein ordinales visuelles Intervall und eine Klassenaufteilung parametrisiert. Ihre Parameterregel lautet demnach:

##### **Regel 5-9** (Parameter der Choroplethentechnik)

**IF**  $\text{Vis}(\text{Choroplethen}(U \rightarrow A), \{U\}, \{A\})$

**THEN ASSERT**  $\text{Param}(\text{Choroplethen}(U \rightarrow A), 1, \text{Intervall}(\text{ordinal}, \text{flächenhaft}))$

**ASSERT**  $\text{Param}(\text{Choroplethen}(U \rightarrow A), 2, \text{Klassenaufteilung}())$

Kreissignaturen sind parametrisiert durch die Farbe und das Größenintervall der Signaturen, sowie durch eine Klassenaufteilung:

##### **Regel 5-10** (Parameter der Kreissignaturtechnik)

**IF**  $\text{Vis}(\text{Kreissignaturen}(U \rightarrow A), \{U\}, \{A\})$

**THEN ASSERT**  $\text{Param}(\text{Kreissignaturen}(U \rightarrow A), 1, \text{Konstante}(\text{Farbe})) ;$

**ASSERT**  $\text{Param}(\text{Kreissignaturen}(U \rightarrow A), 2, \text{Intervall}(\text{Größe})) ;$

**ASSERT**  $\text{Param}(\text{Kreissignaturen}(U \rightarrow A), 3, \text{Klassenaufteilung}())$

Ist eine Visualisierungstechnik primitives Element einer zusammengesetzten Visualisierung, können die verwendeten Kompositionen den Freiheitsgrad der Parameter weiter einschränken. Folglich sind auch für die Kompositionsoperationen, insbesondere für Überlagerung und Animation, Parameterregeln erforderlich. Ist beispielsweise eine flä-

chenförmige Visualisierung  $V_k$  Element einer Überlagerung  $V_1|...|V_k|...|V_n$ , muß sie transparent ausgestaltet werden um die darunterliegenden Signaturen nicht zu überdecken. In der Parameternotation wird dies durch die Aufnahme des einschränkenden Attributes „transparent“ ausgedrückt:

**Regel 5-11** (Parameterwahl bei Überlagerungen flächenförmiger Techniken)

```

IF Vis( $V_1|...|V_k|...|V_n$ ,  $U$ ,  $A$ )
     $\wedge$   $k > 1$ 
     $\wedge$  Param( $V_k$ ,  $i$ ,  $P(a_1, ..., a_{m-1}$ , flächenhaft))
THEN RETRACT Param( $V_k$ ,  $i$ ,  $P(a_1, ..., a_{m-1}$ , flächenhaft)) ;
    ASSERT Param( $V_k$ ,  $i$ ,  $P(a_1, ..., a_{m-1}$ , flächenhaft, transparent))
  
```

Animationen stellen besondere Einschränkungen an die Datenklassifikation. Nur unklassifizierte Techniken sind hier sinnvoll, da mit ihnen ein glatter Übergang zwischen den Bildern erreicht wird. Animationen von Visualisierungen klassifizierter Daten wirken durch die sprunghaften Übergänge zwischen den Einzelbildern irritierend.

**Regel 5-12** (Parameterwahl bei Animationen)

```

IF Vis(Animation( $V$ ),  $U$ ,  $A$ )
     $\wedge$   $V_k$  ist primitives Element von  $V$ 
     $\wedge$  Param( $V_k$ ,  $i$ , Klassenaufteilung( $a_1, ..., a_m$ ))
THEN RETRACT Param( $V_k$ ,  $i$ , Klassenaufteilung( $a_1, ..., a_m$ )) ;
    ASSERT Param( $V_k$ ,  $i$ , Klassenaufteilung(unklassifiziert))
  
```

### 5.6.3 Verwaltung visueller Ressourcen

Nachdem für jede vollständige Visualisierung alle benötigten Parameter und deren Freiheitsgrade ermittelt sind, übernimmt das Ressourcenmanagement die Belegung dieser Parameter mit visuellen Ressourcen. Ressourcen wie Farben und Muster sind knapp: abhängig vom Ausgabemedium stehen nur wenige Farbintervalle oder -konstanten gleichzeitig für eine Visualisierung zur Verfügung. Enthält ein Visualisierungsfenster zu viele Farben, erschwert dies den Überblick und die Interpretation. Viele weitere Constraints sind zu beachten und können in der wachsenden wissenschaftlichen Literatur zu diesem Thema nachgeschlagen werden [CaDi87, Heal96, SDKR95]. Insbesondere Müller geht ausführlich auf die Natur der Darstellungsprimitive Farbe und Textur ein und gibt viele praktische Hinweise für ihren effektiven Einsatz [Müll98].

Das Ressourcenmanagement benötigt also selbst Wissen über visuelle Variablen, über Ausgabemedien und über die visuelle Wahrnehmung. Es muß außerdem die Zuweisung der knappen visuellen Ressourcen optimieren, indem es primitive Visualisierungen mit höherem Gewicht bevorzugt bedient. Zur Realisierung des Ressourcenmanagements ist zunächst eine Tabelle mit Beschreibungen aller vorhandenen visuellen Ressourcen erforderlich. Sie muß pro Ressource folgende Information enthalten:

- Klasse der Ressource: Konstante, selektives, ordinales oder proportionales Intervall
- Dimensionen der Ressource: punktförmig, linienförmig, flächenförmig, volumenförmig (die Dimension der Visualisierungstechnik, auf die sie angewendet wird, muß enthalten sein)
- Priorität der Ressource in ihrer Klasse (Ressourcen, die in ihrer Klasse effektiver sind, erhalten eine höhere Priorität)
- Bereiche des Farb- oder Grauwertspektrums, die von der Ressource belegt werden
- besondere Eigenschaften, wie z.B. Transparenz
- Assoziationen, die mit dieser Ressource verbunden sind (z.B. Blau  $\rightarrow$  kalt, Rot  $\rightarrow$  warm)
- Ausgabemedien, die die Ressource darstellen können
- GIS-Systeme, die die Ressource darstellen können

Liegt eine Beschreibung der verfügbaren Ressourcen nach diesen Charakteristika vor, kann die Vergabe von Ressourcen an eine vollständige Visualisierung  $V$  nach einem relativ einfachen Algorithmus erfolgen:

1.  $R$  sei die Liste aller Ressourcen, die vom aktuellen Ausgabemedium und der aktuellen GIS-Software unterstützt werden und die den Wahrnehmungsfähigkeiten des Benutzers nicht widersprechen.
2. Traversiere die Struktur von  $V$ . Bei Animationen und Überlagerungen verbinde die zusätzlichen Bedingungen mit den Parametern der Operanden.
3. Für jede enthaltene primitive Visualisierung  $V_i$  (in der Reihenfolge des Gewichts) führe Schritt 4 aus:
4. Für jeden Parameter  $P_{i,j} = \text{Param}(V, j, T(a_1, \dots, a_m))$  von  $V_i$ , führe Schritte 5 bis 8 aus:
5.  $R_{i,j} := \{r \in R \mid (r \text{ ist von Typ } T) \wedge (r \text{ erfüllt } a_1, \dots, a_m)\}$
6. Falls Assoziationen zu  $P_{i,j}$  bekannt sind:  
 $T := \{r \in R \mid r \text{ erfüllt die Assoziationen an } P_{i,j}\}$ .  
Ist  $T \neq \emptyset$  dann setze  $R_{i,j} := T$
7. Wähle ein  $r$  aus  $R_{i,j}$  mit maximaler Priorität. Ordne dem Parameter  $P_{i,j}$  den Wert  $r$  zu.
8. Falls  $r$  Farb- oder Grautöne belegt: entferne aus  $R$  alle Ressourcen, die die gleichen oder ähnliche Töne belegen.

Im ersten Schritt des Algorithmus werden aus der Liste der verfügbaren visuellen Ressourcen diejenigen entfernt, die für den aktuellen Kontext nicht effektiv sind. Liegt beispielsweise ein Schwarz/Weiß-Display vor, werden alle Ressourcen, die Farbe enthalten entfernt. Der Algorithmus stellt weiterhin sicher, daß etwaige Assoziationen nach Möglichkeit erfüllt werden. Unter allen Ressourcen, die die Bedingungen an einen Parameter erfüllen, wird immer diejenige mit der höchsten Priorität ausgewählt. Da primitive Visualisierungen mit höherem Gewicht früher behandelt werden, erhalten sie automatisch die effektiveren Ressourcen.

Visuelle Darstellungen unterschiedlicher Datenvariablen sollten vom Betrachter klar zugeordnet werden können. Ihre visuellen Ressourcen müssen deswegen im Wahrnehmungsraum einen gewissen Minimalabstand haben, d.h. visuell klar unterscheidbar sein. Um dies sicherzustellen, entfernt der Algorithmus nach jedem Schritt alle Ressourcen aus der Liste, die der zuletzt zugewiesenen ähnlich sind, also beispielsweise im wahrnehmungsorientierten HSB-Farbraum einen Mindestabstand nicht einhalten.

#### 5.6.4 Datenklassifikation

Für die effektive Klassifikation eines Datensatzes ist zunächst die Funktion der Visualisierung entscheidend. Dient die Visualisierung der explorativen Analyse, dann ist eine möglichst objektive, nur an den Datencharakteristika angelehnte Klassifizierung erforderlich:

Wherever possible it is desirable to select categories that are based on the nature of the distribution to create the most accurate map. Classes in which the values are relatively similar should be chosen. [Tyne92]

Ist beispielsweise bekannt, daß der Datensatz normalverteilt ist, muß auch die Klassen aufteilung dies widerspiegeln. Ist die Verteilungsart jedoch nicht gesichert, kann mit einer numerischen Analyseverfahren eine effektive Klassifizierung gefunden werden. Von manchen Kartographen wird auch die „Inspektion“ des Datensatzes empfohlen. Dazu betrachtet man ein Histogramm der Datenwerte und markiert Häufungen. Je eine oder mehrere Häufungen ergeben dann eine Klasse. Diese Inspektion des Datensatzes kann ebenfalls bis zu einem gewissen Grad durch Expertensystem-Methoden automatisiert werden. Um den Rahmen dieser Arbeit nicht zu sprengen, wird jedoch hierauf nicht im Detail eingegangen. Als Literatur zu diesem Thema ist beispielsweise [CoHo92] zu nennen.

Liegt die Funktion der Visualisierung nicht mehr in der Analyse eines Datensatzes, sondern schon in der Präsentation von Ergebnissen, stellen sich jedoch andere Anforderungen an die Klassifizierung. Jetzt sollen Erkenntnisse über den Datensatz gezeigt und häufig auch bestimmte Aspekte betont werden. Eine Klassifizierung wird dann meistens schon durch die Kommunikationsabsichten vorgegeben und ist nicht mehr automatisch bestimmbar.

## 5.7 Effektivitätsabschätzung

Sind die ersten drei Phasen der Entwurfsmethodik abgeschlossen, wurden bereits alle im Entwurfsraum enthaltenen Visualisierungen der Relation erzeugt und ihre Parameter vollständig festgelegt. Als abschließende Phase evaluiert die Effektivitätsabschätzung diese Visualisierungen und bringt sie in eine eindeutige Reihenfolge bezüglich ihrer Effektivität. Die effektivste Visualisierung kann anschließend automatisch gerendert werden, andere effektive Visualisierungen dem Benutzer auf Anfrage als Alternativen angeboten werden.

Dieser Abschnitt führt zunächst in die Mathematik der Fuzzy-Mengen ein und leitet eine Repräsentationsform für Effektivität basierend auf diesen Mengen her. Anschließend wird gezeigt, wie mit der Fuzzy-Repräsentation und unter Verwendung entsprechender Regeln die Effektivität primitiver und komplexer Visualisierungen ausgewertet werden kann. Als Ergebnis dieser Auswertung steht eine Abschätzung der Visualisierungseffektivität und eine Aussage über die Sicherheit bzw. Unsicherheit der Abschätzung.

### 5.7.1 Fuzzy-Mengen

Fuzzy-Mengen sind eine Erweiterung des klassischen Mengenbegriffs. Sie wurden von Zadeh erstmals in den sechziger Jahren zur Modellierung „unscharfer“ Zusammenhänge vorgeschlagen [Zade65]. Wesentlich später wurde die Idee der Fuzzy-Mengen von der KI-Forschung erneut aufgegriffen und für den Einsatz in regelbasierten Expertensystemen erweitert. Die so entstandene Fuzzy-Logik hat sich in besonderem Maße für die Programmierung von Regelungsanwendungen bewährt [Till92]. Auch in die Geo-Informationssysteme haben Fuzzy-Mengen und Fuzzy-Logik Einzug gehalten. Sie ermöglichen beispielsweise Abfragekriterien an räumliche Datenbanken, die an der natürlichen Sprache angelehnt sind [KoVo91, WaHS90].

Für klassische Mengen  $A$  ist die charakteristische Funktion  $\chi_A$  über die Element-Relation definiert:

$$\begin{aligned}\chi_A(x) &= 0 \Leftrightarrow x \notin A \\ \chi_A(x) &= 1 \Leftrightarrow x \in A\end{aligned}$$

$\chi_A$  kann also nur den Wert 0 oder den Wert 1 annehmen, je nachdem ob  $x$  in der Menge (vollständig) enthalten oder nicht enthalten ist. Zur Definition der Fuzzy-Sets wird dieser Zugehörigkeitsbegriff erweitert:

#### Definition 5-1

Eine *Fuzzy-Menge*  $F$  über einer Grundmenge  $A$  ist eine Menge von geordneten Paaren:

$$F = \{(x, \mu_F(x)) \mid x \in A\}.$$

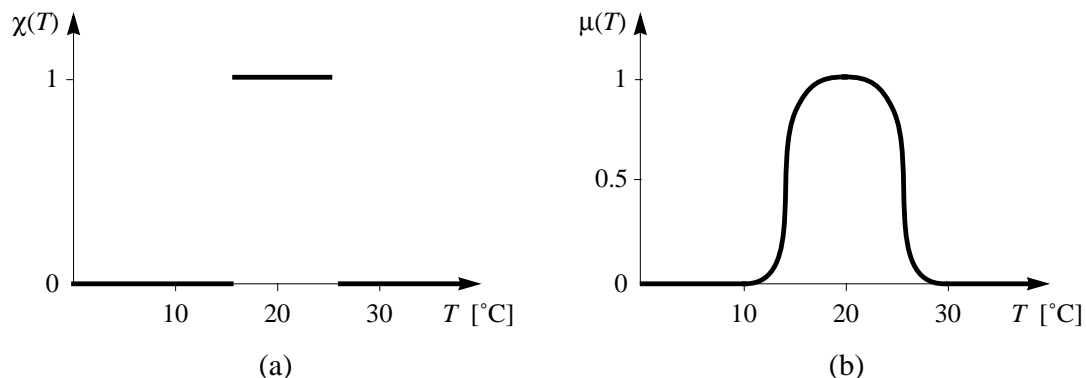
$\mu_F$  ist die *Zugehörigkeitsfunktion* von  $F$ . Sie bildet die Grundmenge  $A$  auf das reelle Intervall  $[0, 1]$  ab:

$$\mu_F : A \rightarrow [0, 1]$$

Der Wert  $\mu_F(x)$  gibt den *Zugehörigkeitsgrad* von  $x$  zu  $F$  wieder. Gilt  $\mu_F(x) = 0$ , ist  $x$  in  $F$  nicht enthalten. Gilt  $\mu_F(x) = 1$ , ist  $x$  in  $F$  vollständig enthalten. Im Falle  $0 < \mu_F(x) < 1$ , ist  $x$  in  $F$  teilweise enthalten und höhere Werte von  $\mu_F(x)$  drücken einen höheren Grad der Zugehörigkeit aus.

### Beispiel 5-6

Fuzzy-Mengen werden u. a. zur Modellierung sogenannter linguistischer Variablen eingesetzt. Eine linguistische Variable gibt die Größe eines Phänomens durch einen an die natürliche Sprache angelehnten Ausdruck wieder. Die Umgebungstemperatur läßt sich z.B. durch Ausdrücke wie „warm“, „heiß“, „mehr oder weniger warm“ wiedergeben. Abb. 5-8 zeigt zwei mögliche Repräsentationen von „warm“. Abb. 5-8a modelliert den Ausdruck durch ein scharfes Intervall mit den Grenzen  $15^\circ\text{C}$  und  $25^\circ\text{C}$ . In unserer Vorstellung des Begriffs warm würden wir jedoch kaum eine scharfe Grenze formulieren und  $15^\circ\text{C}$  als warm,  $14^\circ\text{C}$  jedoch schon als nicht warm bezeichnen. Abb. 5-8b zeigt, wie Fuzzy-Mengen solche Begriffe modellieren: durch unterschiedliche Zugehörigkeitsgrade der verschiedenen Temperaturwerte ergibt sich ein Bild, das der Bedeutung von „warm“ näherkommt.

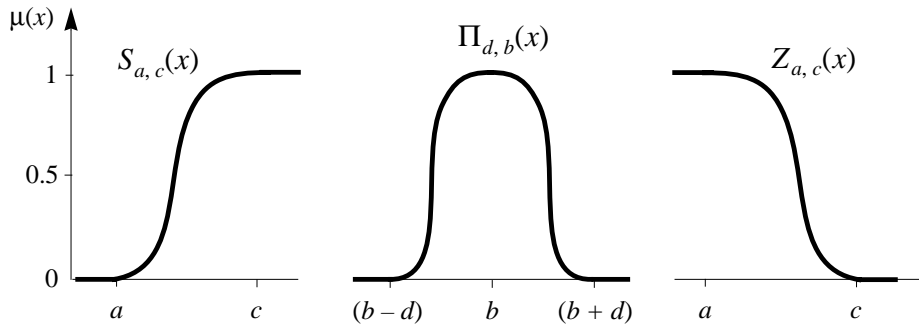


**Abb. 5-8:** Mengenrepräsentationen des Begriffs „warm“: (a) mit einer klassischen Menge und (b) mit einer Fuzzy-Menge

### Standardfunktionen

Der Graph in Abb. 5-8b zeigt eine sogenannte  $\Pi$ -Funktion. Solche Funktionen eignen sich gut als Zugehörigkeitsfunktionen, wenn aus Mengen mit zwei scharfen Intervallgrenzen entsprechende unscharfe Mengen abgeleitet werden sollen. Analog werden sogenannte  $S$ - und  $Z$ -Funktionen definiert, die für die Fuzzyfizierung einer einzelnen Intervallgrenze geeignet sind. Abb. 5-9 zeigt den allgemeinen Graphen für alle drei Funktionen. (Die Namen  $\Pi$ ,  $S$  und  $Z$  geben die Form der Graphen wieder.)





**Abb. 5-9:** Die drei Standard-Zugehörigkeitsfunktionen für Fuzzy-Mengen.

Die  $S$ -Funktion wird durch die Werte  $a$  und  $c$  parametrisiert. Für alle  $x \leq a$  gilt  $S_{a,c}(x) = 0$  und für alle  $x \geq c$  gilt  $S_{a,c}(x) = 1$ . Dazwischen gilt:

$$S_{a,c}(x) = 2\left(\frac{x-a}{c-a}\right)^2, \quad a < x \leq \frac{a+c}{2};$$

$$S_{a,c}(x) = 1 - 2\left(\frac{c-x}{c-a}\right)^2, \quad \frac{a+c}{2} < x \leq c.$$

Die  $Z$ - und die  $\Pi$ -Funktion lassen sich aus  $S_{a,c}(x)$  ableiten:

$$Z_{a,c}(x) = 1 - S_{a,c}(x);$$

$$\Pi_{d,b}(x) = S_{b-d,b}(x), \quad x \leq b;$$

$$\Pi_{d,b}(x) = Z_{b,b+d}(x), \quad x > b.$$

Alle drei Funktionen erreichen ein Maximum von 1. Diese Eigenschaft ist insbesondere in der Fuzzy-Logik zur Normalisierung von Fuzzy-Mengen erwünscht. Tritt eine Zugehörigkeitsfunktion als Ergebnis einer Operation auf und ist sie nicht überall gleich 0, wird sie derart skaliert, daß gilt:

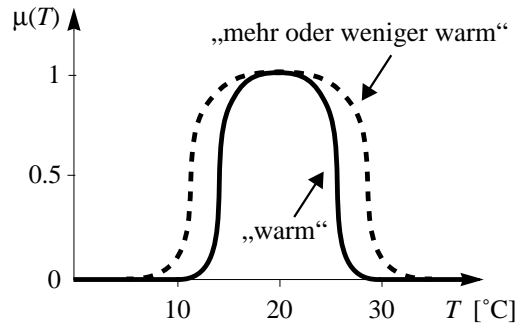
$$\text{Max}\{\mu_F(x) \mid x \in A\} = 1$$

### Fuzzy-Modifikatoren

Eine wichtige Klasse von Operatoren auf Fuzzy-Mengen sind die sogenannten Modifikatoren. Sie verändern die Form der Zugehörigkeitskurve und entsprechend auch die Charakteristika der Fuzzy-Menge. Modifikatoren unterstützen die Modellierung linguistischer Variablen, indem sie die Wirkung von Attributen wie „sehr“, „etwas“, „mehr oder weniger“ auf Fuzzy-Mengen nachbilden. Abb. 5-10 zeigt die Wirkung des Modifikators Dilation auf die Fuzzy-Menge „warm“. Die Dilation  $\text{Dil}(F)$  einer Fuzzy-Menge  $F$  ist über die Zugehörigkeitsfunktion definiert:

$$\mu_{\text{Dil}(F)}(x) = \sqrt{\mu_F(x)}$$

Die Dilation bewirkt eine Aufweichung des Fuzzy-Sets, es wird unschärfer. Im Beispiel der linguistischen Variable *Temperatur* drückt die Dilation das Attribut „mehr oder weniger“ aus.



**Abb. 5-10:** Wirkung des Dilationsoperators auf die Fuzzy-Menge „warm“.

Eine weiterer wichtiger Modifikator ist die Konzentration  $\text{Con}(F)$ , die eine Fuzzy-Menge schärfer macht:

$$\mu_{\text{Con}(F)}(x) = \mu_F(x)^2$$

Mit der Konzentration läßt sich bei vielen Grundmengen das linguistische Attribut „sehr“ modellieren. Nicht jedoch bei der Grundmenge *Temperatur*: die Fuzzy-Menge „sehr warm“ wird besser durch eine Verschiebung von  $\mu_{\text{warm}}$  in positiver Richtung entlang der Temperaturachse modelliert.

### 5.7.2 Wie läßt sich Effektivität repräsentieren?

Wie in Abschnitt 5.1.3 schon angedeutet wurde, ist eine lineare Skala nicht ausreichend, um die komplexen Aspekte der Effektivität von Visualisierungen auszudrücken. Insbesondere widersprüchliche Aspekte und Unsicherheiten erfordern eine Repräsentation in Form einer Fuzzy-Menge.

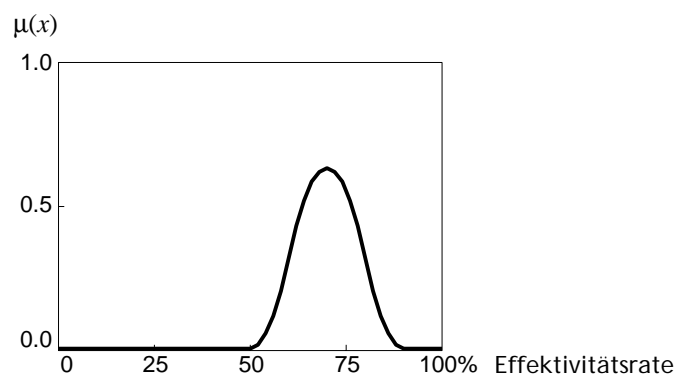
Nehmen wir zunächst eine lineare Skala an, mit einen Wertebereich zwischen 0% („ungenügende“ Effektivität) und 100% („hervorragende“ Effektivität). Verschiedene Grade von Effektivität lassen sich auf dieser Skala problemlos realisieren, einer „guten“ Effektivität kann z.B. der Wert 70% zugeordnet werden, einer „schlechten“ der Wert 30% und einer „mittleren“ der Wert 50%.

Es soll nun die Effektivität von zwei Visualisierungen  $V_1$  und  $V_2$  verglichen werden. Einflußgröße der Effektivität soll in unserem Beispiel die Zielsetzung sein, bestehend aus zwei Zielen  $A$  und  $B$ , jeweils gleich stark gewichtet.  $V_1$  sei bezüglich  $A$  „ungenügend“

und bezüglich  $B$  „hervorragend“.  $V_2$  sei bezüglich beider Ziele „mittel“. Eine Auswertung der Gesamteffektivität ergibt für beide Visualisierungen einen Wert von 50%, also „mittel“. Die Gesamteffektivität beider Visualisierungen kann also auf der linearen Skala nicht unterschieden werden, obwohl sie doch in Wirklichkeit sehr unterschiedliche Effektivitätscharakteristika aufweisen.

Auch wenn die Effektivitätsbewertung von  $V_2$  im Gegensatz zu der von  $V_1$  umstritten ist, d.h. nicht allgemeingültig festgelegt werden kann, läßt sich dies in der linearen Skala nicht widerspiegeln. Die Festlegung auf einen „scharfen“ Wert erlaubt keine Repräsentation von Unsicherheiten.

Beide Mängel lassen sich durch die Verwendung von Fuzzy-Mengen als Repräsentationsform für Effektivität beseitigen. Abb. 5-11 zeigt eine solche Fuzzy-Repräsentation einer „guten“ Effektivität. Sie wird nun nicht mehr durch einen scharfen Wert, sondern durch eine Zugehörigkeitsfunktion dargestellt. Deren Kurve erreicht ihr Maximum bei einer Rate von 70% und ist bei Raten unter 50 % bzw. über 90% gleich Null. Unsicherheiten lassen sich hier durch flachere Kurvenverläufe ausdrücken. Widersprüchliche Effektivitäten äußern sich in komplexeren Kurvenverläufen mit mehreren lokalen Maxima.



**Abb. 5-11:** Fuzzy-Repräsentation einer „guten“ Effektivität

An der Zugehörigkeitsfunktion dieser Fuzzy-Menge fällt auf, daß ihr Maximum nicht der Wert 1, sondern ein Wert kleiner als 1 ist. Im Gegensatz zu der Fuzzy-Menge „warm“ in Abb. 5-8 ist sie also nicht Maximum-normiert. Der Grund hierfür ist, daß für das Arbeiten mit Fuzzy-Effektivitäten eine Normierung des Integrals der Zugehörigkeitsfunktion, also der Fläche unter der Kurve, besser geeignet ist. Für flächennormierte Fuzzy-Mengen läßt sich die gewichtete Summe, eine wichtige Operation auf Effektivitäten, sehr einfach definieren (siehe Abschnitt 5.7.4).

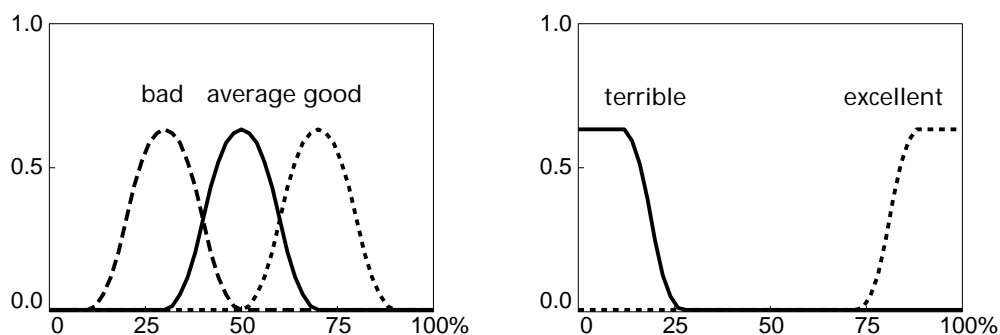
### 5.7.3 Syntax und Fuzzy-Repräsentation der Effektivität

Bevor die Fuzzy-Repräsentation im Detail vorgestellt wird, definieren wir zunächst eine einfache Syntax für die linguistische Variable „Effektivität“. Sie umfaßt fünf Hauptgrade

der Effektivität: terrible, bad, average, good und excellent. Bad und good können durch die Zusätze very oder slightly modifiziert (verstärkt bzw. abgeschwächt) werden. Eine Aussage über die relative Sicherheit bzw. Unsicherheit des Effektivitätsgrades erreicht man durch die Zusätze certainly und probably.

$$\begin{aligned}
 \text{Effectiveness} &= [ \text{Certainty} ] \text{ Value} \\
 \text{Certainty} &= \{ \text{certainly} \mid \text{probably} \} \\
 \text{Value} &= \{ \text{PlainValue} \mid \text{ModValue} \} \\
 \text{PlainValue} &= \{ \text{average} \mid \text{terrible} \mid \text{excellent} \} \\
 \text{ModValue} &= [ \text{Modification} ] \{ \text{good} \mid \text{bad} \} \\
 \text{Modification} &= \{ \text{very} \mid \text{slightly} \}
 \end{aligned}$$

Abb. 5-12 zeigt die Fuzzy-Repräsentanten der fünf Hauptgrade von Effektivität. Bad, average und good sind durch ähnliche  $\Pi$ -Funktion dargestellt, die an verschiedenen Stellen ihr Maximum erreichen. Terrible und excellent werden mit einer Z- bzw. S-Funktion dargestellt.



**Abb. 5-12:** Die fünf Hauptgrade von Effektivität

Eine Repräsentation der Fuzzy-Modifikatoren ist in Abb. 5-13 zu sehen. Slightly und very bewirken eine geringe Verschiebung der Kurve nach unten bzw. nach oben entlang der horizontalen Achse. Probably bewirkt eine Dilation, d.h. die Kurve wird gestreckt und die Fuzzy-Menge damit unschärfer. Analog dazu bewirkt certainly eine Konzentration, die Kurve wird gestaucht und die Fuzzy-Menge schärfer.

## 5.7.4 Rechnen mit Effektivitäten

### Gewichtete Summen

Die wichtigste arithmetische Operation auf Effektivitäten ist die gewichtete Summe. Die Effektivität einer Visualisierung wird im allgemeinen durch eine ganze Reihe von Faktoren oder Bedingungen beeinflusst. Diese Bedingungen üben einen unterschiedlich starken Einfluß auf die Gesamteffektivität aus. Ein Beispiel für solche unterschiedlich gewichte-

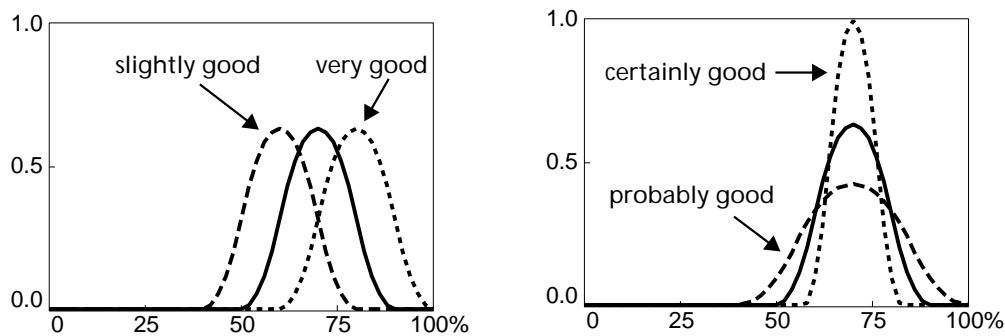


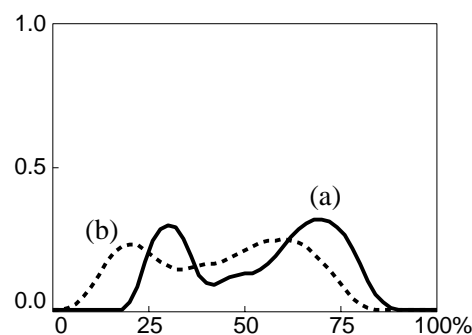
Abb. 5-13: Fuzzy-Modifikationen von „good“

ten Bedingungen sind mehrfache Visualisierungsaufgaben, wie sie in Abschnitt 4.2.4 definiert wurden. Um eine Gesamteffektivität zu errechnen, müssen die Teileffektivitäten bezüglich der unterschiedlichen Interpretationsaufgaben unter Einbeziehung ihres Gewichtes addiert werden.

Die gewichtete Summe der Fuzzy-Effektivitäten  $E_i$  ist über die gewichtete Summe ihrer Zugehörigkeitsfunktionen  $\mu_i$  definiert:

$$E = \sum_i w_i E_i \Leftrightarrow \mu_E(x) = \sum_i w_i \mu_i(x)$$

Abb. 5-14a zeigt ein Beispiel für eine solche gewichtete Summe von Effektivitäten.



**Abb. 5-14:** (a) gewichtete Summe  $S = 0.3$  certainly bad +  $0.2$  average +  $0.5$  good, (b) ihre Modifikation slightly less probable (worse ( $S$ )).

Durch die Flächennormierung

$$\int_0^{100} \mu_i(x) dx = c$$

mit konstantem Wert  $c$  wird gewährleistet, daß jede Fuzzy-Menge entsprechend ihres Gewichts in die Summe einfließt. Wären die Summanden nicht Flächen- sondern Maximum-normiert, würden nicht nur die Gewichte, sondern auch die unterschiedlichen Flächeninhalte der Zugehörigkeitsfunktionen ihren Anteil an der Summe bestimmen. Die Gewichtung der Summe wäre somit verfälscht.

Gilt für die Gewichte  $\sum_i w_i = 1$ , dann ist auch die gewichtete Summe wieder flächennormiert, wie aus der Additionsregel für Integrale folgt.

### Modifikationen

Neben der gewichteten Addition stellen auch die Modifikationen eine wichtige Operation auf Effektivitäten dar. So kann es in der regelbasierten Auswertung gelegentlich vorkommen, daß vorher getroffene Abschätzungen nach „unten“ oder „oben“ korrigiert werden müssen, da einige der Standardannahmen in den Regeln nicht zutrafen. Gleichermäßen kann eine Korrektur der Sicherheit einer Effektivitätsbewertung erforderlich sein.

Modifikationen werden als einstellige Operationen auf Fuzzy-Effektivitäten definiert:

$$\begin{aligned} \text{ModEff} &= [ \text{ModProb} ] [ \text{Shift} ] ( \text{Effectiveness} ) \\ \text{ModProb} &= [ \text{much} \mid \text{slightly} ] \{ \text{worse} \mid \text{better} \} \\ \text{Shift} &= [ \text{much} \mid \text{slightly} ] \{ \text{less certain} \mid \text{more certain} \} \end{aligned}$$

Abb. 5-14b zeigt ein Beispiel für eine modifizierte Effektivität. Die Operatoren müssen so beschaffen sein, daß auch die Ergebnisse wieder flächennormiert sind.

### 5.7.5 Wie läßt sich Effektivität abschätzen?

In den vorangehenden Abschnitten wurde die Frage nach einer geeigneten Repräsentation von Effektivität beantwortet. Es bleibt zu klären, wie die Effektivität von Visualisierungen in der abschließenden Phase des Visualisierungsentwurfs abgeschätzt und verglichen werden kann.

Die Effektivität einer Visualisierung hängt zunächst davon ab, wie gut die visuellen Variablen die Datensatzvariablen wiedergeben. Regeln über die Effektivität visueller Variablen finden sich in der Literatur zur graphischen Semiologie und zur Visualisierungsmethodik (siehe Abschnitt 2.5.2), die Effektivitätscharakteristika der visuellen Variablen wurden experimentell bestätigt [CIMc84]. Primitive Visualisierungstechniken setzen im allgemeinen mehrere visuelle Variablen zusammen mit einer spezifischen Ausgestaltung und Methodik ein, um Datenvariablen darzustellen. Solche Techniken, insbesondere Formen thematischer Karten und statistischer Diagramme, verkörpern implizit schon Wissen über die Effektivität visueller Variablen, z.B. fordert die Choroplethentechnik ein Intervall aus einer ordinalen oder proportionalen visuellen Variablen, da nur diese für quantitative Daten effektiv sind. Unter den visuellen Variablen, die diese Forderung erfüllen, wird in der Parameteroptimierungsphase schließlich diejenige ausgewählt, die im gegebenen

Kontext optimal ist. Damit ist die Effektivität von Visualisierungen auf der fundamentalen Ebene der visuellen Variablen sichergestellt.

Auf einer höheren Ebene variiert die Effektivität der Techniken jedoch stark mit der Zielsetzung der Visualisierung. Betrachtet man z.B. die Aufgaben einer Visualisierung, kann eine Choroplethenkarte desselben Datensatzes einerseits sehr effektiv (für das Erkennen räumlicher Verteilungen) aber auch uneffektiv (für das Ablesen einzelner Werte) sein. Regelwerke und Lehrbücher über Visualisierungsmethodik geben über diese verschiedenen Verwendungsarten der Techniken zumindest qualitative Aussagen. Detaillierte Effektivitätsmaße für die wichtigsten Techniken bezüglich verschiedener Aufgaben konnten im Rahmen dieser Arbeit durch eine Reihe von Tests gewonnen werden (siehe Kapitel 6).

Als weitere Faktoren für die Bewertung der Visualisierungseffektivität sind die Charakteristika der Benutzer und der Ausgabemedien zu beachten. Hier läßt sich jedoch festhalten, daß diese Elemente des Visualisierungskontexts innerhalb einer Visualisierungssituation typischerweise wesentlich weniger variieren werden als die Interpretationsabsichten oder andere Aspekte der Zielsetzung. Dies schlägt sich auch in der Diskussion der Visualisierungseffektivität in Lehrwerken nieder: gewöhnlich wird zunächst von individuellen Unterschieden abstrahiert und die Effektivität für einen „Standard-“ Benutzer und das „typische“ Ausgabemedium (in der Kartographie die Papierkarte und in der technisch-wissenschaftlichen Visualisierung der Graphikbildschirm) betrachtet. Auf die Berücksichtigung von abweichenden Benutzern, anderen Medien und anderen Umgebungen wird dann in einer gesonderten Diskussion eingegangen.

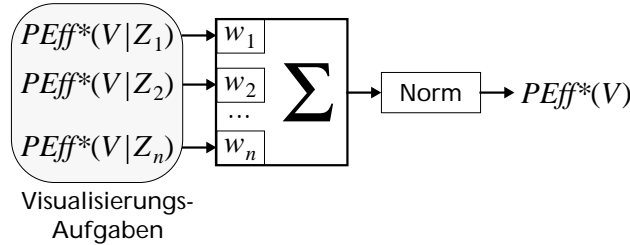
Ein vergleichbares Vorgehen ermöglicht auch in der automatisierten Auswertung von Visualisierungseffektivität, alle entscheidenden Faktoren zu berücksichtigen. Dazu wird, ausgehend von den Interpretationsabsichten und anderen Aspekten der Zielsetzung zunächst eine *Standardeffektivität* der Visualisierung ermittelt. Dazu stehen Regeln zur Verfügung, die die Visualisierungseffektivität in einem *Standardkontext* bewerten. Dieser Kontext beschreibt einen Standardbenutzer, ein Standardausgabemedium und eine standardmäßige Softwareumgebung. Beim Standardbenutzer wird man z.B. davon ausgehen, daß er über eine durchschnittliche visuelle Wahrnehmung verfügt und insbesondere nicht farbenblind ist. Das Standardausgabemedium eines GIS zur Analyse und Planung ist der Farbmonitor. Charakteristika, die von diesem Standardkontext abweichen, in dieser Arbeit *Ereignisse* genannt, werden durch gesonderte Regeln behandelt. Diese Ereignisregeln versuchen zu erfassen, wie die abweichende Situation die Standardeffektivität der Visualisierung verändert [Jung96b].

### 5.7.6 Primitive Effektivitätsabschätzung

#### **Standardeffektivität**

Am Anfang der Auswertung einer komplexen Visualisierung steht die Ermittlung der Standardeffektivität aller enthaltenen primitiven Visualisierungen. Dazu stehen für jede der primitiven Techniken Regeln zur Verfügung, die ihre Effektivität bezüglich der verschiedenen Visualisierungsaufgaben bewerten. Benutzer können mehrere Aufgaben

gleichzeitig anfordern und diese auch gewichten. Zur Ermittlung der Standardeffektivität werden die entsprechenden Teileffektivitäten gewichtet summiert (siehe Abb. 5-15).



**Abb. 5-15:** Berechnung der Standardeffektivität

Auf eine primitive Visualisierung  $V$  treffen genau diejenigen Aufgaben  $Z_i$  zu, die eine oder mehrere abhängige Variablen von  $V$  referenzieren. Die Standardeffektivität  $PEff^*(V|Z_i)$  von  $V$  bezüglich  $Z_i$  wird aus der entsprechenden Regel ermittelt und mit dem Gewicht  $w_i$  von  $Z_i$  skaliert. Eine Summation für alle zutreffenden Visualisierungsaufgaben ergibt die Standardeffektivität  $PEff^*(V)$  von  $V$ :

$$PEff^*(V) = \text{Norm}\left(\sum_{i \text{ zutreffend}} w_i PEff^*(V|Z_i)\right)$$

Eine Normalisierung ist an dieser Stelle erforderlich, da es im allgemeinen auch Aufgaben gibt, die nicht auf  $Z_i$  zutreffen und deswegen  $\sum_{i \text{ zutreffend}} w_i < 1$  sein kann.

### Beispiel 5-7

Die folgenden Regeln bestimmen die Standardeffektivität der Choroplethentechnik. Die Effektivitätswerte, die diese Regeln zuweisen, sind im Rahmen dieser Arbeit empirisch ermittelt worden (siehe Kapitel 6). Die folgende Regel behandelt den Fall einer Choroplethenkarte bei variablenlokalen Aufgaben:

#### **Regel 5-13** (Standardeffektivität der Choroplethentechnik, variablenlokal)

```

IF Task( $Z_i$ ) = read THEN  $PEff^*(Choro|Z_i) := \text{bad}$ 
ELSE IF Task( $Z_i$ ) = find THEN  $PEff^*(Choro|Z_i) := \text{slightly good}$ 
ELSE IF Task( $Z_i$ ) = compare THEN  $PEff^*(Choro|Z_i) := \text{certainly bad}$ 
ELSE IF Task( $Z_i$ ) = trend THEN  $PEff^*(Choro|Z_i) := \text{good}$ 
  
```

### Standardeffektivität bei variablenübergreifenden Aufgaben

Bei der Ermittlung der Effektivität von Aufgaben bezüglich mehrerer Variablen  $A_1, \dots, A_m$  muß zusätzlich berücksichtigt werden, wie die Choroplethenkarte für  $V_1$  mit anderen Karten kombiniert wird, um alle Variablen auszudrücken. Hier werden vier Fälle unterschieden:



- *Gleich/Gleich*: Die Choroplethenkarte für  $V_1$  wird mit weiteren Choroplethenkarten überlagert.
- *Gleich/Verschieden*: Die Choroplethenkarte wird mit weiteren Karten überlagert, darunter auch andere Techniken.
- *Gleich + Gleich*: Es wird eine Sequenz aus Choroplethenkarten verwendet.
- *Gleich + Verschieden*: Es wird eine Sequenz aus Choroplethenkarten und anderen Techniken verwendet

### Beispiel 5-8

In der folgenden Regel wird zwischen diesen vier Fällen unterschieden, um die Standardeffektivität der Choroplethentechnik bezüglich der Aufgabe *correlate* zu bestimmen. Es zeigt sich, daß die Choroplethenkarten nur in Kombination mit weiteren Choroplethenkarten für das Aufzeigen einer Korrelation geeignet sind. (In dieser Regel sind die zugewiesenen Werte durch *probably* modifiziert, da sie nicht direkt gemessen, sondern aus der Literatur hergeleitet wurden [Monm90].)

**Regel 5-14** (Standardeffektivität der Choroplethentechnik für *correlate*)

```

IF Gleich/Gleich THEN  $PEff^*(Choro|Z_i) := \text{probably good}$ 
ELSE IF Gleich/Verschieden THEN  $PEff^*(Choro|Z_i) := \text{probably bad}$ 
ELSE IF Gleich + Gleich THEN  $PEff^*(Choro|Z_i) := \text{probably slightly good}$ 
ELSE IF Gleich + Verschieden THEN  $PEff^*(Choro|Z_i) := \text{probably bad}$ 

```

Für die primitive Visualisierung *Choroplethen(Bezirk  $\rightarrow$  Temperatur)* soll nun die Standardeffektivität bestimmt werden. Wir übernehmen die Visualisierungszielsetzung aus Beispiel 4-2:

```

( objectives
  ( interpretation 25% compare Niederschlag )
  ( interpretation 35% find Temperatur )
  ( interpretation 40% correlate Niederschlag Temperatur
                                domains Monat )
)

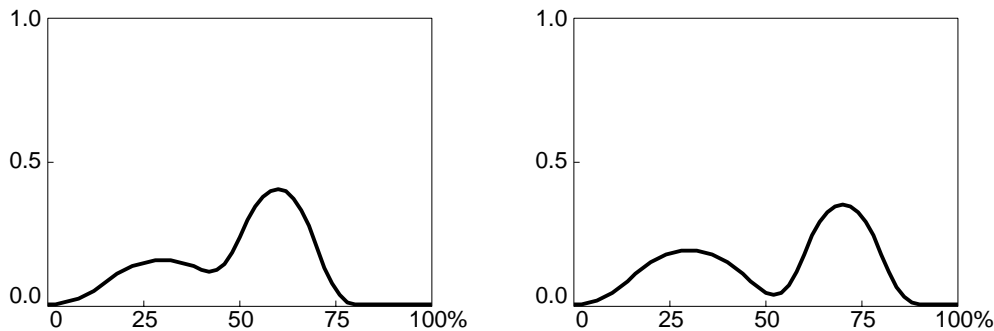
```

Von den Visualisierungsaufgaben treffen die zweite und die dritte auf *Choroplethen(Bezirk  $\rightarrow$  Temperatur)* zu, da sie die abhängige Variable *Temperatur* referenzieren. Für die Standardeffektivität der Choroplethenkarte (bei Überlagerung einer Kreissignaturenkarte) ergibt sich der folgende Wert:

Norm(0.35 (slightly good) + 0.2 (probably bad)).

Für *Kreissignaturen(Bezirk  $\rightarrow$  Niederschlag)* wird die Standardeffektivität analog bestimmt. In den Effektivitätsregeln zur Kreissignaturentechnik (die hier nicht aufgeführt werden sollen) wird folgender Wert zugewiesen:

Norm( $0.25$  (good) +  $0.2$  (probably bad)).



**Abb. 5-16:** Standardeffektivitäten von Choroplethen(*Bezirk*  $\rightarrow$  *Temperatur*) und von Kreissignaturen(*Bezirk*  $\rightarrow$  *Niederschlag*)

### Ereignisse und primitive Effektivität

Ist die Standardeffektivität einer primitiven Visualisierung bestimmt, muß im nächsten Schritt die Erkennung und Behandlung von Ereignissen erfolgen. Als Ereignis werden alle Bedingungen bezeichnet, die sich von den Voraussetzungen der Standardeffektivitätsregeln signifikant unterscheiden. Dazu zählen natürlich in erster Linie die Abweichungen vom Standardkontext.

Abweichungen vom Standardkontext werden allerdings schon in früheren Phasen der Visualisierungsmethodik berücksichtigt. Hier ist insbesondere die Parameteroptimierung und der Algorithmus zur Vergabe visueller Ressourcen zu nennen. Schritt 1 dieses Algorithmus (siehe Seite 125) läßt als verfügbare Ressourcen nur solche zu, die im aktuellen Kontext effektiv sind. Dies kann jedoch zur Folge haben, daß die einer primitiven Visualisierung tatsächlich zugewiesene visuelle Variable weniger effektiv ist, als im Standardfall zu erwarten wäre. Die Standardeffektivität muß entsprechend (negativ) korrigiert werden. Abweichungen vom Standardkontext wirken sich also zunächst indirekt, über die Vergabe visueller Ressourcen, auf die Visualisierungseffektivität aus.

Es können aber auch Korrekturen der Standardeffektivität einer Visualisierungstechnik erforderlich sein, wenn sie die gestellte Aufgabe zwar primär erfüllt, die Semantik des Datensatzes aber nicht optimal repräsentiert. Dies kann zu mißverständlichen Interpretationen führen und sollte deswegen in der primitiven Effektivität berücksichtigt werden.

Die folgende Regel behandelt einen solchen Fall. Sie korrigiert die Standardeffektivität der Kreissignaturen nach unten, wenn die dargestellte Variable nicht aus Absolutwerten (wie durch die proportional großen Kreissignaturen suggeriert) sondern aus Dichten oder Verhältniszahlen besteht.

**Regel 5-15** (Ereignis: Kreissignaturen bei Dichten und Verhältniszahlen)

**IF**  $\text{Vis}(\text{Kreissignaturen}(U \rightarrow A), \dots)$

$\wedge A$  besteht aus Dichten oder Verhältniszahlen

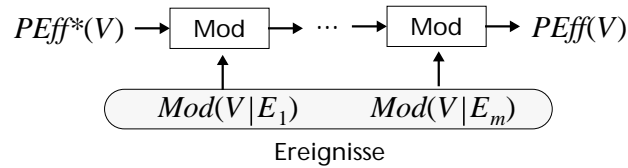
**THEN**  $\text{Mod}(V|E_k) = \text{slightly less certain (slightly worse)}$

Die in solchen Regeln zugewiesenen Korrekturwerte sind eher heuristischer Natur. Sie drücken eine Tendenz aus, beruhen aber nicht wie die Standardeffektivitäten auf Meßergebnissen. Sie schließen deshalb in der Regel auch eine Modifikation der Sicherheit des Wertes ein (slightly less certain).

Treffen für eine primitive Visualisierung  $m$  Ereignisse  $E_1, \dots, E_m$  zu, werden die Modifikationen  $\text{Mod}(V|E_k)$  bezüglich dieser Ereignisse sukzessive auf die Standardeffektivität von  $V$  angewendet. Es ergibt sich schließlich die Gesamteffektivität  $\text{PEff}(V)$  der primitiven Visualisierung  $V$ :

$$\text{PEff}(V) = \text{Mod}(V|E_1) \bullet \dots \bullet \text{Mod}(V|E_m) (\text{Norm}(\text{PEff}^*(V)))$$

Abb. 5-17 zeigt das Schema der Errechnung der primitiven Effektivität aus Standardeffektivität und Ereignissen.



**Abb. 5-17:** Berechnung der primitiven Effektivität

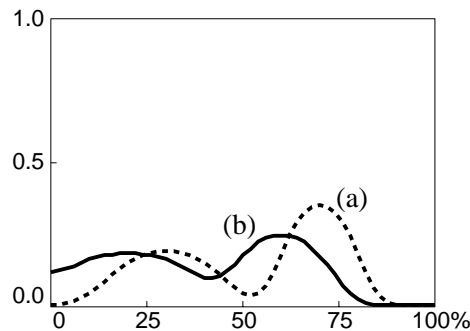
### Beispiel 5-9

Wir betrachten die primitive Visualisierung  $\text{Kreissignaturen}(\text{Bezirk} \rightarrow \text{Niederschlag})$ . Die Variable *Niederschlag* ist bezüglich der Fläche normiert, es handelt sich also um Dichtewerte und das Ereignis 5-15 trifft zu. Unter der Annahme, daß keine weiteren Ereignisse zutreffen, ergibt sich eine Modifikation von slightly less certain (slightly worse). Insgesamt ergibt sich für die primitive Effektivität von  $\text{Kreissignaturen}(\text{Bezirk} \rightarrow \text{Niederschlag})$  folgender Wert (siehe auch Abb. 5-18a):

slightly less certain ( slightly worse ( Norm(0.25 (good) + 0.2 (probably bad))))).

### 5.7.7 Komplexe Effektivitätsabschätzung

Die Abschätzung der Effektivität einer vollständigen Visualisierung  $V$  erfolgt in ähnlichen Stufen wie die primitive Abschätzung: zunächst werden die Effektivitäten der in  $V$  enthaltenen primitiven Visualisierungen gemäß ihrem Gewicht summiert. Anschließend



**Abb. 5-18:** Primitive Effektivität (a) und Standardeffektivität (b) von Kreissignaturen (*Bezirk* → *Niederschlag*).

werden alle zutreffenden Ereignisse ermittelt und die entsprechenden Modifikationen ausgeführt.

Hier stellt sich zunächst die Frage, wie die primitiven Visualisierungen gewichtet werden sollen. Per Definition der Ausdrucksfähigkeit ist jede abhängige Variable des Datensatzes in genau einer primitiven Visualisierung dargestellt. (Primitive Visualisierungen können mehrere abhängige Variablen darstellen.) Das Gewicht  $u_j$  einer primitiven Visualisierung  $V_j$  ergibt sich somit aus der Summe der Gewichte aller in  $V_j$  enthaltenen abhängigen Variablen. (Zum Gewicht der Variablen siehe Abschnitt 4.2.4 auf Seite 91).

### Struktur-Ereignisse

Auch auf der Ebene der komplexen Visualisierungen können Ereignisse die Gesamteffektivität im positiven oder im negativen Sinn beeinflussen. Für den Standardfall einer idealen komplexen Visualisierung müssen unter anderem folgende Aussagen zutreffen:

- Visualisierungen sollten möglichst nicht viele unterschiedliche Techniken verwenden (dies erschwert das Verständnis und die Interpretation)
- Visualisierungen sollen nur eine kleine Anzahl von Fenstern haben, wenn sie den Vergleich von Variablen unterstützen sollen
- Animationssequenzen sollten nicht zu kurz sein
- die Bedeckung von Karten mit Symbolen sollte nicht zu groß sein (Übersichtlichkeit)

Im hier vorgestellten Ereigniskonzept lassen sich solche Zusammenhänge durch Ereignisregeln formulieren, die eine Modifikation der Effektivität je nach dem Grad der Abweichung vom Standardfall vornehmen. Dazu wird eine Rampenfunktion  $Rampe(a_{Min} \rightarrow M_{Min}, a_{Max} \rightarrow M_{Max}, a)$  definiert, die für eine Größe  $a$  (z.B. die Anzahl der Frames einer Animationssequenz) eine Modifikation berechnet, die zwischen einem geringen Wert von  $M_{Min}$  für  $a_{Min}$  und dem maximalen Wert von  $M_{Max}$  liegt.

Die folgenden vier Regeln sind beispielhafte Implementierungen unter Verwendung einer solchen Rampenfunktion:

**Regel 5-16** (Ereignis: mehrere Windows)

**IF** die Aufgabenstellung enthält variablenübergreifende Aufgaben  
 $\wedge$   $a$  ist die Anzahl der Windows von  $V$   
 $\wedge$   $a \geq 5$

**THEN**  $Mod(V|E) := Rampe(5 \rightarrow \text{slightly worse}, 10 \rightarrow \text{much worse}, a)$

**Regel 5-17** (Ereignis: kurze Animationssequenzen)

**IF**  $V = Anim(\dots)$   
 $\wedge$   $a$  ist die Anzahl der Frames von  $V$   
 $\wedge$   $a < 25$

**THEN**  $Mod(V|E) := Rampe(25 \rightarrow \text{slightly worse}, 1 \rightarrow \text{much worse}, a)$

**Regel 5-18** (Ereignis: mehrfache Techniken)

**IF**  $a$  ist die Anzahl der Techniken von  $V$   
 $\wedge$   $a \geq 2$

**THEN**  $Mod(V|E) := Rampe(2 \rightarrow \text{slightly worse}, 4 \rightarrow \text{much worse}, a)$

**Regel 5-19** (Ereignis: Symbolbedeckung)

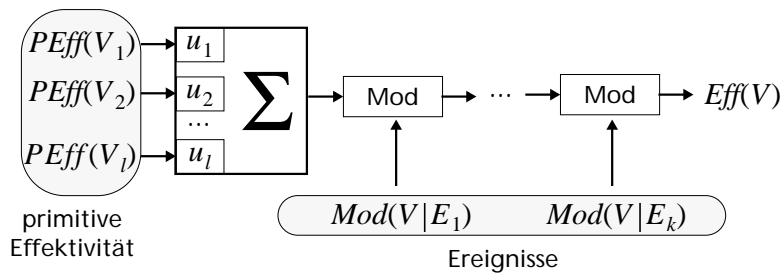
**IF**  $c$  ist der prozentuelle Grad der Symbolbedeckung von  $V$   
 $\wedge$   $c \geq 50\%$

**THEN**  $Mod(V|E) := Rampe(50\% \rightarrow \text{slightly worse}, 100\% \rightarrow \text{much worse}, c)$

### Gesamteffektivität

Treffen die Ereignisse  $E_1$  bis  $E_k$  zu, ergibt sich insgesamt die folgende Gleichung für die komplexe Effektivität einer vollständigen Visualisierung  $V$  (siehe auch Abb. 5-19):

$$Eff(V) = Mod(V|E_1) \cdot \dots \cdot Mod(V|E_k) \left( \sum_j u_j PEff(V_j) \right)$$



**Abb. 5-19:** Abschätzung der komplexen Effektivität

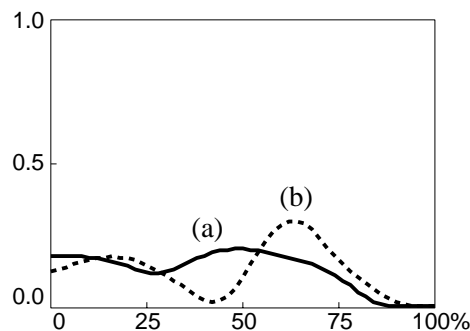
**Beispiel 5-10**

$V = \text{Animation}(\text{Choroplethen}(\text{Bezirk} \rightarrow \text{Temperatur}) \mid \text{Kreissignaturen}(\text{Bezirk} \rightarrow \text{Niederschlag}), \text{Monat})$  ist eine vollständige Visualisierung der Relation Klima. Die Effektivitäten der zwei primitiven Elemente  $V_1 = \text{Choroplethen}(\text{Bezirk} \rightarrow \text{Temperatur})$  und  $V_2 = \text{Kreissignaturen}(\text{Bezirk} \rightarrow \text{Niederschlag})$  wurden bereits in den Beispielen 5-7 bzw. 5-9 berechnet.

Für  $V_1$  ergibt sich ein Gewicht von 0.55, für  $V_2$  ein Gewicht von 0.45, entsprechend den abhängigen Variablen, die sie darstellen. Der primitive Anteil der Effektivität von  $V$  ergibt sich also wie folgt:

$$Eff_{\text{Prim}} = 0.55 \text{ Norm}(PEff(V_1)) + 0.45 \text{ Norm}(PEff(V_2))$$

Es treffen die Ereignisse 5-17 und 5-18 zu. Die Gesamteffektivität der Visualisierung  $V$  ist in Abb. 5-20 dargestellt (Ein wesentliches ausführliches Beispiel zur Effektivitätsbewertung, bei dem die Fuzzy-Effektivitäten unterschiedlicher Entwürfe verglichen und erläutert werden, wird noch in Abschnitt 7.3.2 gegeben.)



**Abb. 5-20:** Gesamteffektivität (a) und primitiver Anteil (b) der Beispielvisualisierung

### 5.7.8 Defuzzifizierung

Arbeitet man mit Fuzzy-Mengen, ist es an bestimmten Stellen notwendig, aus der Fuzzy-Menge einen scharfen Wert abzuleiten. In der Fuzzy-Logik spricht man dabei von Defuzzifizierung. Beim Visualisierungsentwurf sind es im wesentlichen zwei Stellen, an denen aus den Fuzzy-Effektivitäten konkrete, scharfe Aussagen abgeleitet werden müssen. Zum einen ist es notwendig, alle ausdrucksfähigen Visualisierungen gemäß ihrer Effektivität zu ordnen, um den effektivsten Entwurf zu finden und um alternative Entwürfe in der richtigen Reihenfolge zu präsentieren. Zum anderen sollen Benutzer bei Erklärungen nicht mit der Fuzzy-Effektivität konfrontiert werden. Statt dessen sollte sie textuell dargestellt werden.

### Ordnungskriterien

Als Ordnungskriterium bietet sich zunächst der Schwerpunkt (*CoG*, center of gravity) der Fuzzy-Menge an. Er entspricht in etwa dem gewichteten Mittel der in die Fuzzy-Menge eingeflossenen Summanden und gibt damit die durchschnittliche Effektivität wieder. Der Schwerpunkt einer Fuzzy-Menge ist wie folgt definiert:

$$CoG(E) = \frac{\int_0^{100} x \mu_E(x) dx}{\int_0^{100} \mu_E(x) dx}$$

Neben der durchschnittlichen Effektivität als Hauptmaß sind noch zwei weitere Eigenschaften der Fuzzy-Effektivität von Bedeutung:

- *die Varianz:* Ist sie hoch, dann ist die Effektivität relativ unscharf und der Durchschnittswert kann nicht als gesichert angenommen werden. Unscharfe Gesamteffektivitäten entstehen durch einen hohen Anteil unscharfer Teileffektivitäten, durch entsprechende Modifikationen (less certain) oder durch die Addition sehr gegensätzlicher Teileffektivitäten.
- *der Anteil niedriger Effektivitätsgrade:* Ist er hoch, dann hat die Visualisierung einen hohen Anteil an ineffektiven Aspekten. Selbst wenn ihre durchschnittliche Effektivität akzeptabel ist, sollte eine Visualisierung mit hohem Anteil niedriger Effektivitätsgrade abgewertet werden.

Die Definition einer Varianz für Fuzzy-Mengen kann an die statistische Varianz angelehnt werden. Sei  $CoG(E)$  der Schwerpunkt einer Fuzzy-Effektivität  $E$ , dann ist ihre Varianz definiert als:

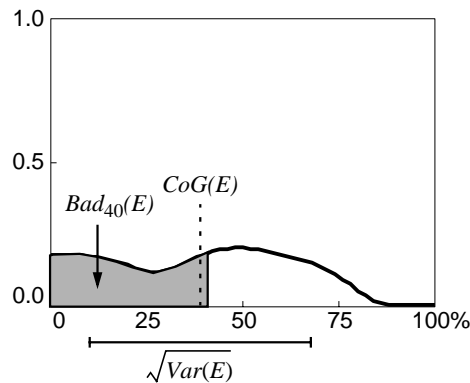
$$Var(E) = \int_0^{100} (x - CoG(E))^2 \mu_E(x) dx$$

Der Anteil niedriger Effektivitätsgrade läßt sich unter Verwendung einer Schranke  $s$  definieren. Er ist der Anteil der Gesamtfläche, der unter  $s$  liegt:

$$Bad_s(E) = \frac{\int_0^s \mu_E(x) dx}{\int_0^{100} \mu_E(x) dx}$$

### Priorisierung

Alle vier Kriterien, die Abb 5-21 noch einmal anhand der in Beispiel 5-10 berechneten Gesamteffektivität zeigt, müssen gleichzeitig beachtet werden. Höchste Priorität erhält dabei sicherlich der Schwerpunkt. Die Prioritäten der anderen Kriterien sind weniger of-



**Abb. 5-21:** Ordnungskriterien für Fuzzy-Effektivitäten

fensichtlich und hängen teilweise vom Kontext der Visualisierung ab. Bei der explorativen Analyse eines Datensatzes können ungesicherte oder unkonventionelle Visualisierungen hilfreich sein, da sie vielfältige Sichten des Datensatzes ermöglichen. In diesem Kontext führt also eine relativ hohe Varianz der Effektivität nicht zwangsläufig zur Abwertung. Bei Präsentationsgraphiken andererseits muß sie niedrig sein, da nur gesicherte Visualisierungen publiziert werden sollten. Ein niedriger Anteil uneffektiver Aspekte sollte jedoch für alle Visualisierungen angestrebt werden.

Für die explorative Analyse ist eine beispielhafte Priorisierung der Ordnungskriterien in der folgenden Liste zusammengefaßt:

1.  $CoG(E)$  maximal
2.  $Bad_s(E)$  minimal
3.  $Var(E)$  minimal

### Textuelle Darstellung

Eine textuelle Darstellung der Fuzzy-Effektivität einer Visualisierung, die zur Information des Benutzers gedacht ist, sollte auf relativ kurzem Raum die wichtigsten Aspekte der Effektivität zusammenfassen. Als wichtigste Aspekte sind wieder die drei Ordnungskriterien zu nennen:

- *der Schwerpunkt:* Er wird durch den linguistischen Ausdruck für die am nächsten liegende Effektivitätsrate benannt.
- *die Varianz:* Sie wird zunächst in wenige Stufen diskretisiert und die unterschiedlichen Varianzgrade werden dann durch Schlüsselwörter ausgedrückt. Nimmt man beispielsweise fünf Stufen an, ergeben sich folgende Varianzgrade: very roughly, roughly, (keine Nennung), precisely, very precisely.
- *der Anteil niedriger Effektivität:* Auch er wird diskretisiert und in verschiedenen Graden wiedergegeben. (keine Nennung), slightly ineffective, ineffective, very ineffective.



**Beispiel 5-11**

Für die Effektivität der vollständige Visualisierung aus Beispiel 5-10 ergeben sich folgende Werte aus der Defuzzifizierung:

- $CoG(E) \approx 38.7$
- $\sqrt{Var(E)} \approx 57.8$
- $Bad_{40}(E) \approx 50.1$
- $Low(E) = 0$

Basierend auf diese Werte lautet die textuelle Darstellung der Effektivität:

very roughly: slightly bad  
(very ineffective)

## 5.8 Erweiterungen der Methodik

Um die Darstellung der Entwurfsmethodik in Kapitel 5 überschaubar und verständlich zu halten, mußten einige einschränkende Annahmen getroffen werden. In einzelnen wurde bisher angenommen, daß Datensätze aus einer einzigen Relation bestehen und daß der Entwurfsraum für vollständige Visualisierungen immer relativ klein bleibt. In diesem Abschnitt soll gezeigt werden, wie diese Einschränkungen durch einfache Erweiterungen der Methodik wegfallen können.

### 5.8.1 Kompakte Visualisierung mehrerer Relationen

Die in Abschnitt 5.5 vorgestellten Regeln und Operatoren sind bisher nur zur Komposition von Visualisierungen geeignet, die die gleiche Menge  $U$  von unabhängigen Variablen ausdrücken. So können z.B. Visualisierungen von  $Bezirk \rightarrow Temperatur$  und  $Bezirk \rightarrow Niederschlag$  zusammengesetzt werden. Diese Einschränkung rührte daher, daß mit Überlagerung und Sequenz nur Teilrelationen aus einer vorhergegangenen funktionalen Dekomposition zusammengesetzt werden sollten, um unkonventionelle Entwürfe zu vermeiden.

Es ist aber vielfach notwendig, auch Visualisierungen strukturell unterschiedlicher Relationen zu überlagern. Nehmen wir z.B. an, daß die Niederschlagsmengen nicht wie bisher als Durchschnittswerte für jeden Bezirk und aggregiert für jeden Monat vorliegen. Statt dessen betrachten wir eine Relation  $R_2$ , die die exakte Niederschlagsmenge an einer Reihe von Meßorten für jeden Tag des Jahres enthält. Der Datensatz besteht also aus zwei Relationen:

$$R_1: \text{Bezirk, Monat} \rightarrow \text{Temperatur}$$

$$R_2: \text{Ort, Datum} \rightarrow \text{Niederschlag}$$

Auch für diesen komplexen Datensatz läßt sich eine Visualisierung analog zu der aus den vorhergehenden Beispielen vorstellen: eine Überlagerung einer Karte der Niederschlagsmengen über einer Karte der Temperaturen. Da die Niederschlagsmengen nun nicht mehr für Bezirke aggregiert (flächenförmig) sondern an Orten (punktförmig) vorliegen, können sie nicht mit Choroplethen, aber beispielsweise mit Kreissignaturen dargestellt werden. Eine Voraussetzung für die Überlagerung der beiden Karten wäre, daß sich die dargestellten geographischen Bereiche und auch die zulässigen Maßstäbe der Darstellungen überlappen.

Auch eine Animation dieser Überlagerung macht Sinn: Zeiteinheit der Animation ist nicht mehr ein Monat sondern ein Tag. Die Darstellung von  $R_2$  wird in der Animation an jedem Tag mit neuen Meßwerten aktualisiert, die Darstellung von  $R_1$  lediglich am Ersten des Monats, wenn also neue Werte für *Temperatur* vorliegen. Insgesamt ist diese Animation jedoch nur zulässig, wenn sich die Zeitabschnitte überlappen, zu denen für die beiden Relationen Daten vorliegen.

Animation und Überlagerung von Teilvisualisierungen sind also auch dann zulässig, wenn sie unterschiedliche unabhängige Variablen  $U_1$ ,  $U_2$  ausdrücken. Die Variablen  $U_1$  und  $U_2$  müssen dann jedoch derselben *Ordnungsdimension* angehören und ihre Wertebereiche müssen sich überlappen. Gehören zwei Variablen derselben Ordnungsdimension an, können ihre Werte direkt miteinander verglichen und eindeutig angeordnet werden. (Eine lineare Anordnung wird jedoch nicht gefordert.) Einige Beispiele von Ordnungsdimensionen sind:

- *der geographische (2D-) Raum:* Variablen wie *Bezirk*, *Ort*, *Straße* können (über ihre geographische Lage) verglichen und angeordnet werden.
- *der 3D-Raum* (analog, über die 3D-Lage)
- *die Zeit:* Variablen wie *Monat*, *Datum*, *Uhrzeit* können verglichen und angeordnet werden.
- *ähnliche Index-Mengen*

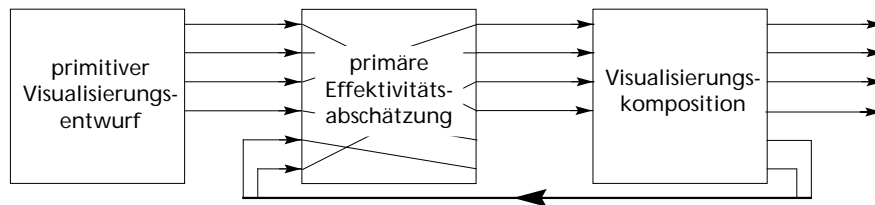
Soll die Visualisierungsmethodik auch Datensätze mit mehrfachen Relationen abdecken, müssen also die Kompositionsregeln entsprechend verändert werden. Die Kompositionsoperationen fügen dann nicht mehr ausschließlich solche Teilrelationen zusammen, die aus einer vorhergehenden Dekomposition entstanden sind.

### 5.8.2 Einschränkung des Suchbaums

Die dargestellte Entwurfsstrategie bedient sich zur Bestimmung ausdrucksfähiger und effektiver Visualisierungen der erschöpfenden Suche. Es wird ein Suchbaum aufgebaut, der alle im Entwurfsraum vorhandenen Visualisierungen des Datensatzes enthält. Dies stellt bei einer beschränkten Anzahl primitiver Techniken und Kompositionsoperatoren kein Problem dar und ermöglicht eine vollständige Untersuchung aller Visualisierungsmög-

lichkeiten. Wächst jedoch die Anzahl primitiver Techniken, die vom System unterstützt werden, an oder werden weitere Kompositionsooperatoren eingeführt, kann der Suchraum sehr groß werden. Dann muß die Methodik so verändert werden, daß ineffektive Visualisierungen frühzeitig aussortiert werden und der Suchraum entsprechend schrumpft.

Dazu kann die Effektivitätsabschätzung in zwei Phasen, eine primäre und eine sekundäre Abschätzung, aufgespalten werden. Die primäre Abschätzung erfolgt vor der Kompositionsphase und entfernt alle partiellen Visualisierungen aus dem Suchraum, die eine niedrige Standardeffektivität besitzen. Es werden also nur Visualisierungen mit einem Mindestmaß „potentieller“ Effektivität zusammengesetzt (siehe Abb. 5-22). Durch die Wahl eines Schwellwertes für die Standardeffektivität läßt sich der Suchraum entsprechend verkleinern. Nach abgeschlossener Komposition und Parameteroptimierung erfolgt dann eine sekundäre Abschätzung, in der die Effektivität der Visualisierungen endgültig bewertet wird.



**Abb. 5-22:** Die ersten drei Phasen einer modifizierten Entwurfsmethodik.

Dieses heuristische Suchverfahren wird Beam-Search genannt [BaFe81, Mack91]. Unter bestimmten Voraussetzungen ermöglicht es die frühzeitige Erkennung von vernachlässigbaren Kandidaten und damit eine effektive Beschneidung des Suchbaums. In der Entwurfsmethodik für Visualisierungen sind diese Voraussetzungen erfüllt, wenn die Standardeffektivitätsabschätzung nicht zu stark von der endgültigen Effektivitätsabschätzung abweicht.

## 5.9 Zusammenfassung und Diskussion

### Zusammenfassung der Ergebnisse

Kapitel 5 stellt eine Methodik für den automatisierten Visualisierungsentwurf in Geo-Informationssystemen vor. Basierend auf das Modell zur Charakterisierung von Daten und Visualisierungskontext sowie auf einem Regelsatz zur Ausdrucksfähigkeit und Effektivität primitiver Visualisierungstechniken werden Visualisierungen raumbezogener Datensätze in vier Phasen entworfen:

1. *Primitiver Visualisierungsentwurf:* Der Datensatz wird mit den Ausdrucksregeln aller verfügbaren Visualisierungstechniken verglichen. Wenn eine Teilrela-

tion des Datensatzes die Ausdruckskriterien einer Technik erfüllt, werden Visualisierungen erzeugt.

2. *Visualisierungskomposition:* Die primitiven Visualisierungen werden durch Kompositionsoperationen wie Überlagerung und Animation wiederholt zusammengesetzt bis alle vollständigen Visualisierungen des Datensatzes gefunden sind.
3. *Visualisierungsparameter-Optimierung:* Die vollständigen Visualisierungen werden traversiert und alle offenen Parameter wie Signaturen, Farben und Klassenaufteilungen bestimmt. Unter Verwendung entsprechender Regeln werden die Parameter für jede vollständige Visualisierung optimiert und zugewiesen.
4. *Effektivitätsabschätzung:* Die vollständigen Visualisierungen werden erneut traversiert und die Effektivität der enthaltenen primitiven Visualisierungen sowie die Effektivität der durchgeführten Kompositionen abgeschätzt. Danach lassen sich die Visualisierungen bezüglich ihrer Effektivität in eine eindeutige Reihenfolge bringen.

In der Abschätzungsphase wird die Effektivität der Visualisierungen explizit durch Fuzzy-Mengen modelliert. Die Abschätzung stützt sich auf die gewichtete Summation empirisch bestimmbarer Effektivitätswerte für die verschiedenen Techniken und Visualisierungsaufgaben in einem Standardkontext. Abweichungen vom jeweiligen Standardkontext lassen sich durch heuristische Regeln berücksichtigen, die die Standardwerte entsprechend korrigieren.

Das Arbeiten mit Fuzzy-Mengen ermöglicht insbesondere

- die Modellierung widersprüchlicher Aspekte in der Effektivität einer Visualisierung,
- eine angemessene Unterscheidung zwischen gesichertem und ungesichertem Wissen über Effektivität, und
- den Einsatz heuristischer (Ereignis-) Regeln, die Abweichungen vom Standardkontext erfassen und modellieren.

## Diskussion

Die erste und die zweite Phase der Methodik basieren auf der bahnbrechenden Arbeit von Mackinlay [Mack86a] und erweitern bzw. adaptieren das von ihm entwickelte Verfahren für die 2D-Visualisierung von Geodaten. Auf dem Gebiet der automatisierten Symbolisierung thematischer Karten wird die Arbeit von Zhang und Buttenfield [ZhBu95] durch die neu entwickelte Methodik so ergänzt, daß die Beschränkung auf univariate Datensätze entfällt. Die Fuzzy-Effektivität modifiziert und verallgemeinert die in der Visualisierungshilfe VH [GLLH97] errechneten Eignungsvektoren für Visualisierungsmethoden derart, daß auch Abweichungen vom Standardkontext der Visualisierung berücksichtigt werden können.

Als neue und signifikante Aspekte der in dieser Arbeit entwickelten Entwurfsmethodik für GIS-Visualisierungen sind hervorzuheben:

- Die Low-Level-Operatoren, die Mackinlay zur Komposition graphischer Primitive beschreibt, werden durch eine Reihe von High-Level-Operatoren zur Komposition von 2D-Visualisierungen von Geodaten ersetzt.
- Visualisierungseffektivität wird erstmals explizit repräsentiert und abgeschätzt. Die Repräsentation erfolgt mittels Fuzzy-Mengen und spezifischer Summations- und Auswertungsoperationen.
- Durch die Fuzzy-Modellierung der Visualisierungseffektivität lassen sich Zielsetzung, Benutzer und Ausgabemedium im Visualisierungsentwurf berücksichtigen und über das Ereignis-Konzept vielfältige Effektivitätsregeln integrieren.

Die Verwendung der High-Level-Operatoren zur Visualisierungskomposition ist vorteilhaft für den Entwurf von Geodaten-Visualisierungen, da sie besser geeignet sind, die rigiden Verfahren der thematischen Kartographie zu implementieren. Durch den Wegfall der *Mark Composition* werden unkonventionelle Darstellungen fast vollständig vermieden. Dies bedeutet jedoch auch, daß der Entwurfsraum für Visualisierungen eingeschränkt ist und die insbesondere für das Stadium der explorativen Analyse erwünschte Zahl vielfältiger und kontroverser Sichten des Datensatzes beschränkt wird. Vorteilhaft an den High-Level-Operatoren ist jedoch auch, daß sie sich von der gängigen GIS-Software bereits in gewissem Maße direkt ausführen lassen.

Insbesondere die Fuzzy-Modellierung und -abschätzung der Visualisierungseffektivität läßt sich auch über den GIS-Kontext hinaus im automatisierten Visualisierungsentwurf einsetzen. Überall dort, wo sich Visualisierungsaufgaben konkret benennen und sich die Effektivität der Verfahren abschätzen oder gar theoretisch bestimmen läßt, können mit dieser Methodik Visualisierungen bewertet werden. So zum Beispiel in der technisch-wissenschaftlichen Visualisierung, in der Analyse von Multiparameterdaten oder auch im Data Mining. Einschränkend muß zur Fuzzy-Abschätzung der Effektivität bemerkt werden, daß sie zwar die Formulierung sehr vielfältiger Effektivitätsregeln zuläßt, die Genauigkeit der Abschätzung jedoch abnimmt, je weiter man sich von den Standardfällen der Visualisierungen entfernt. (In Kapitel 7 wird die Genauigkeit der Abschätzung anhand eines Beispiels im Detail besprochen.)

Einen unverzichtbaren Bestandteil intelligenter Visualisierungssysteme kann die vorgestellte Methodik jedoch nicht ersetzen: eine umfassende Regelbasis. In diesem Kapitel wurden Regeln zur GIS-Visualisierung nur für einen kleinen aber wichtigen Teil der Visualisierungstechniken (die quantitativ-thematischen Karten) und einige exemplarische Aspekte des Visualisierungskontextes aufgestellt. Sie bilden einen Grundstock an Regeln für die intelligente GIS-Visualisierung, müssen jedoch in einem schwierigen und arbeitsintensiven Prozeß der Wissensakquisition ergänzt werden. Für die vorgestellte GIS-Visualisierungsmethodik wird diese Wissensakquisition in den Jahren 1998-2000 im ESPRIT-Projekt CommonGIS mit Förderung der Europäischen Kommission durchgeführt (siehe Abschnitt 8.3.2).

**Ausblick**

Die Methodik zum automatisierten Entwurf von GIS-Visualisierungen läßt sich in mehreren Richtungen weiter ausbauen und vertiefen. Als wichtigste Gebiete weiterer Forschung können identifiziert werden:

- die Entwicklung von Operatoren für die 3D-Komposition und damit die Unterstützung von 3D-Visualisierungstechniken für Geodaten, und
- die Integration von Verfahren zur kartographischen Generalisierung (als Vorphase der primitiven Visualisierung) und damit die Gewährung einer gewissen Maßstabsunabhängigkeit.

# Kapitel 6

## Empirische Bestimmung der Effektivität von GIS-Visualisierungsformen

In der Auswertungsphase greift die Entwurfsmethodik auf Regeln zu, die die Effektivität von Visualisierungsentwürfen bezüglich des aktuellen Kontextes bewerten. Die Regelbasis muß also Wissen über die relative Effektivität aller angebotenen Visualisierungsformen in allen möglichen Ausprägungen des Kontexts enthalten. Insbesondere muß bekannt sein, welche Visualisierungsform für eine bestimmte Aufgabenstellung die effektivste ist und um welchen Grad ineffektiver die anderen Formen sind. In der Literatur werden die Visualisierungsformen aber überwiegend nur qualitativ bewertet [z.B. Selg90b, KeKe92]. In den bekannten quantitativen Untersuchungen zur Eignung von Visualisierungsmethoden [CIMc84] wird nicht auf die Vielzahl verschiedener Aufgaben eingegangen, die eine Visualisierung unterstützen kann.

Da also in der Literatur keine ausreichend detaillierten Regeln über den Eignungsgrad kartographischer Visualisierungsformen vorliegen, wurden im Rahmen dieser Arbeit solche Werte empirisch ermittelt. Es wurden eine Reihe von Tests durchgeführt, die die relative Eignung einiger Visualisierungsformen für verschiedene Aufgaben einer Visualisierung messen. Ziel dieser empirischen Untersuchung war zum einen die Gewinnung realistischer Standardeffektivitätswerte für die Regeln der Beispielimplementierung VIZARD. Zum anderen sollte mit diesen Tests gezeigt werden, daß auch bei den kartographischen Visualisierungsformen signifikante Unterschiede bezüglich ihrer Eignung für verschiedene Aufgaben bestehen.

Die Tests wurden konzipiert in Anlehnung an Clevelands Untersuchung von Balken- und Liniendiagrammen [CIMc84] sowie an neuere Tests, die die Effektivität von Methoden und Geräten der Computergraphik untersuchen [HaDo96, Jaco94, Volb96, Ware88]. Im Rahmen dieser Arbeit konnten nicht alle Visualisierungsformen und -aufgaben betrachtet werden, die für GIS in Frage kommen. Um den Aufwand vertretbar zu halten, wurden die vier wichtigsten Visualisierungsformen für die vier elementaren Aufgaben getestet.

## 6.1 Konzeption der Effektivitätstests

### 6.1.1 Hypothesen

Folgende beiden Hypothesen wurden aufgestellt:

- $H_1$ : Zur Lösung einer gegebenen Visualisierungsaufgabe sind die verschiedenen zulässigen Visualisierungsformen unterschiedlich gut geeignet.
- $H_2$ : Für verschiedene Visualisierungsaufgaben ergeben sich unterschiedliche Reihenfolgen der Eignung der einzelnen Visualisierungsformen.

Diese Hypothesen sollen sich auf die vier elementaren Visualisierungsaufgaben,

- *Werte Ablesen*,
- *Werte Finden*,
- *Werte Vergleichen* und
- *Trend Erkennen*

und vier wichtige quantitativ-kartographische Visualisierungsformen,

- Balkendiagrammkarten,
- Kreissignaturenkarten,
- univariate Choroplethenkarten mit Symbolisierung durch Helligkeit und Farbsättigung (kurz: Choroplethen/Intensität) und
- univariate Choroplethenkarten mit Symbolisierung durch ein Punkteraster (kurz: Choroplethen/Raster),

erstrecken. Bei diesen Visualisierungsformen wird jeweils ein kontinuierlicher Signaturverlauf vorausgesetzt, d.h. es findet keine Klasseneinteilung der Wertebereiche statt.

### 6.1.2 Testdesign

#### Schema

Das Grundelement im Testdesign ist der Einzeltest. Darin wird einer Testperson  $P$  die Testaufgabe  $A$  vorgelegt, die sie anhand der Darstellung eines Datensatzes  $D(A, i)$  durch die Visualisierungsform  $V$  beantworten soll. Die Darstellung wird der Person nur für eine begrenzte Zeit gezeigt. Aus der Antwort der Testperson und der tatsächlichen Lösung wird mit dem aufgabenspezifischen Fehlermaß dann der Fehler  $F(P, A, V, i)$  bestimmt.

Dieser Einzeltest wird nun mit derselben Testperson und derselben Kombination aus  $A$  und  $V$  wiederholt, allerdings werden jeweils unterschiedliche Datensätze  $D(A, 1)$ ,



$D(A, 2), \dots, D(A, 10)$  dargestellt. Diese 10 Wiederholungen eines Einzeltests bilden eine Testsequenz (kurz: Sequenz).

Jede Person führt pro Aufgabe vier Sequenzen durch, je eine für die vier verschiedenen Visualisierungsformen. Alle Sequenzen zu einer Aufgabe  $A$  verwenden dieselben Datensätze  $D(A, 1), \dots, D(A, 10)$ , allerdings ist jeweils die Reihenfolge permutiert. Pro Person ergeben sich für die vier Aufgaben Werte *Ablezen* (**A**), *Werte Finden* (**F**), *Werte Vergleichen* (**V**) und *Trend Erkennen* (**T**) sowie die vier Visualisierungsformen Balkendiagramm (**1**), Kreissignaturen (**2**), Choroplethen/Intensität (**3**) und Choroplethen/Raster(**4**) 16 Sequenzen und damit 160 Einzeltests von jeweils einigen Sekunden Dauer (Tabelle 6-1).

Personen	Aufgabe A		Aufgabe F		Aufgabe V		Aufgabe T	
$P_1$	VF		VF		VF		VF	
	1		1		1		1	
	2		2		2		2	
	3		3		3		3	
	4		4		4		4	
$P_2$	VF		VF		VF		VF	
	1		1		1		1	
	2		2		2		2	
	3		3		3		3	
	4		4		4		4	
...	...		...		...		...	

**Tabelle 6-1:** Das Testschema. Jedes der kleinen Kästchen stellt einen Einzeltest dar. (VF = Visualisierungsform)

Die verschiedenen Aufgaben erfordern naturgemäß unterschiedliche Fehlermaße. Es können deswegen nur die Fehlerwerte aus Einzeltests zur gleichen Aufgabe verglichen und statistisch ausgewertet werden. Das gesamte Experiment zerfällt deswegen in vier Teile, entsprechend den vier Spalten unter Aufgabe A bis Aufgabe T in Tabelle 6-1. Diese vier Teile müssen getrennt statistisch ausgewertet werden.

### Charakterisierung

Jeder Teil des Experiments enthält eine unabhängige Variable, die Visualisierungsform, mit den Klassen **1** bis **4** und eine abhängige Variable, den Fehlerwert. Der Test folgt dem Within-Subject- bzw. dem Randomized-Block-Design, d.h. jede Person wird bezüglich aller 4 Visualisierungsformen getestet. Dies hat gegenüber den Between-Subject-Designs den Vorteil, daß Unterschiede zwischen den Testpersonen besser von den Unterschieden zwischen den Visualisierungsformen getrennt und damit statistisch signifikantere Ergebnisse erzielt werden [Lane93]. Allerdings muß bei den Within-Subject-Designs ausgeschlossen werden, daß das Testergebnis einer Person für eine Visualisierungsform  $V_2$  da-

durch beeinflußt wird, daß die Person schon vorher mit der Visualisierungsform  $V_1$  getestet wurde (sog. Carryover-Effekte).

### Störfaktoren

Neben der abhängigen Variablen Visualisierungsform ist zu erwarten, daß noch weitere Variablen das Testverhalten und damit die erreichten Fehlerwerte beeinflussen. Diese Störfaktoren gilt es zu minimieren. Folgende Störfaktoren wurden bei der Konzeption der Tests beachtet:

- *Bearbeitungszeit:* Alle Testbilder einer Sequenz werden für eine feste, kurze Zeitspanne gezeigt. Dadurch werden sowohl individuelle Unterschiede zwischen der Bearbeitungsdauer verschiedener Personen als auch bei der Bearbeitungsdauer der Aufgaben zu den verschiedenen Visualisierungsformen eliminiert.
- *Vorkenntnisse:* Alle Testpersonen verfügen über die gleiche Information bezüglich der Tests und es bestehen keine relevanten Unterschiede in den Vorkenntnissen im Bereich kartographische Visualisierung.
- *Displayhardware:* Alle Testpersonen führen die Tests nacheinander am selben Displaygerät aus (die Displayhardware beeinflußt im allgemeinen die Visualisierungseffektivität).
- *Carryover-Effekte:* Symmetrische Effekte können durch Counterbalancing minimiert werden, d.h. verschiedene Personen führen die Tests in unterschiedlicher Reihenfolge aus.

## 6.2 Testaufbau und Umsetzung

Testfälle, Aufgabe und Fehlermaß müssen so definiert sein, daß (a) in jeder Sequenz im Durchschnitt einige Fehler produziert werden, (b) in jeder Sequenz nicht zu viele Fehler produziert werden und (c) sich Unterschiede im durchschnittlichen Fehler ergeben. Dazu wurden die Testfälle und die zugrundeliegenden Testdaten in Vortests überprüft und modifiziert.

Tabelle 6-2 zeigt diejenigen Sequenzen, die konzipiert und auch durchgeführt wurden. Die Sequenz **V4** wurde zwar konzipiert, nach den Vortests jedoch gestrichen, da sie im Vergleich zu anderen Sequenzen derselben Aufgabe bei sonst identischen Bedingungen extrem hohe Fehlerraten erzielte.

Alle Sequenzen und Einzeltests folgen demselben Schema. Zu Beginn einer Sequenz wird die Testperson über die zu lösende Aufgabe unterrichtet, kann sich anhand einiger Versuchsbilder über die Visualisierungsform informieren und die Aufgabe probieren. Ist die Person bereit, drückt sie einen Startknopf und es wird für genau fünf Sekunden (bei der Aufgabe *Trend Erkennen* für sieben Sekunden) ein Testbild gezeigt. Anschließend beantwortet die Testperson eine einfache Frage, die sich direkt auf die Aufgabe bezieht, z.B.

	<i>Werte Ablesen</i>	<i>Werte Finden</i>	<i>Werte Vergleichen</i>	<i>Trend Erkennen</i>
Balkendiagramme	<b>A1</b>	<b>F1</b>	<b>V1</b>	<b>T1</b>
Kreissignaturen	<b>A2</b>	<b>F2</b>	<b>V2</b>	<b>T2</b>
Choroplethen/Intensität	<b>A3</b>	<b>F3</b>	<b>V3</b>	<b>T3</b>
Choroplethen/Raster	<b>A4</b>	<b>F4</b>		<b>T4</b>

**Tabelle 6-2:** Überblick über die Testsequenzen (V4 wurde nicht durchgeführt).

„welcher Wert wurde für das Land Hessen dargestellt?“. Die Frage ist für alle Sequenzen einer Aufgabe identisch.

Alle Tests wurden mit 14 Testpersonen aus einer klar definierten Gruppe (mit Informatik-Hochschulausbildung im Alter von 25-35 Jahren) durchgeführt. Dadurch konnten mit vertretbarem Aufwand Effektivitätsmaße gewonnen werden, die zwar nicht repräsentativ, aber doch aussagekräftig für ein typisches GIS-Applikationsszenario sind. Für die 15 Sequenzen wurden pro Testperson zwischen 30 und 45 Minuten benötigt. (Danach wurden in weiteren Sequenzen noch einige weitere Kombinationen von Visualisierungsformen und Aufgaben getestet, auf die hier allerdings nicht eingegangen wird.)

Die Tests wurden in Form von HTML-Files umgesetzt und mit Netscape, einem verbreiteten HTML-Browser durchgeführt. Diese Plattform wurde gewählt, da so die Test-Interfaces schnell entwickelt werden konnten und die Browser sehr stabil und allen Testpersonen vertraut sind. Alle Testbilder wurden mit VIZARD, also dem Zielsystem, erzeugt. Die Parameter der Visualisierungsformen wurden jeweils für die Aufgaben optimiert. Als geographische Basisdaten dienten die Länder- und Kreisgrenzen der Bundesrepublik, da sie einerseits typisch für GIS-Anwendungen sind und andererseits den Testpersonen vertraut waren. Die Testdatensätze wurden mit speziell abgestimmten Rechteck- und Normalverteilungen erzeugt.

### 6.2.1 Testaufbau zu Werte Ablesen

Abb. 6-1 zeigt den Aufbau für die Testaufgabe *Werte Ablesen*. Die Testperson sieht eine thematische Karte der Bundesländer für 5 Sekunden und muß den für Hessen dargestellten Wert ablesen. Die Testpersonen wurden dazu angehalten, den Wert möglichst genau abzulesen und beispielsweise nicht auf ein Vielfaches von 5 zu runden. Die Abbildungen 6-2 bis 6-5 zeigen typische Testbilder für die fünf getesteten Visualisierungsformen.

Die Testdaten sind rechteckverteilt in den Grenzen  $[0,100]$ . Jede der Sequenzen **A1** – **A4** besteht aus 10 Einzeltests, die die 10 verfügbaren Datensätze in zufälliger Reihenfolge zeigen. Als Fehlermaß dient in den A-Tests die Differenz zwischen dem abgelesenen und dem tatsächlich dargestellten Wert.

Testreihe A1, Nr. 1  
(Werte ablesen)

Instruktionen:

- Nachdem Sie auf START gedrückt haben, erscheint für einige Sekunden eine Karte.
- Versuchen Sie auf dieser Karte den dargestellten Wert für Hessen möglichst genau abzulesen.
- Anschließend beantworten sie die Frage und drücken OK.

**START**

Ein paar unverbindliche Versuche gefällig?  
[Versuch 1] [Versuch 2] [Versuch 3]

Frage:  
Welcher Wert wurde für Hessen angezeigt?

OK

**Abb. 6-1:** Testaufgabe A: *Werte Ablesen*

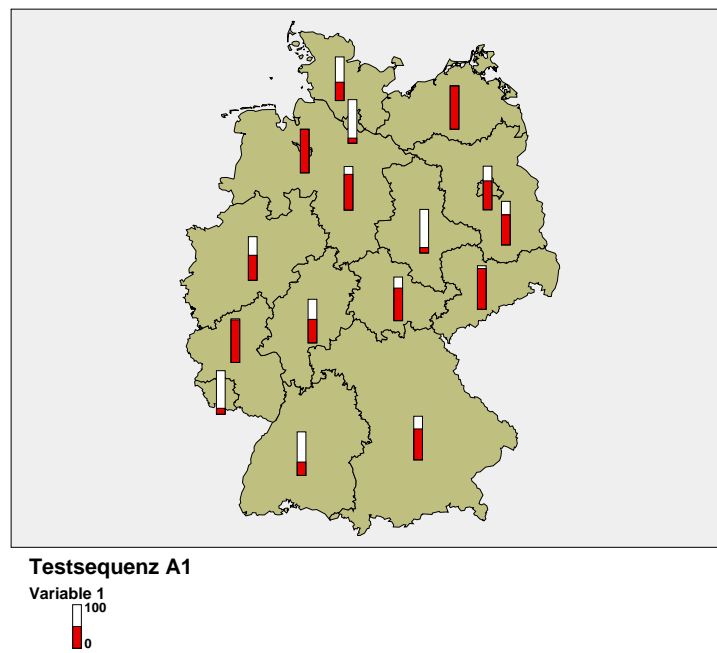


Abb. 6-2: Ein Testbild zur Balkendiagrammkarte

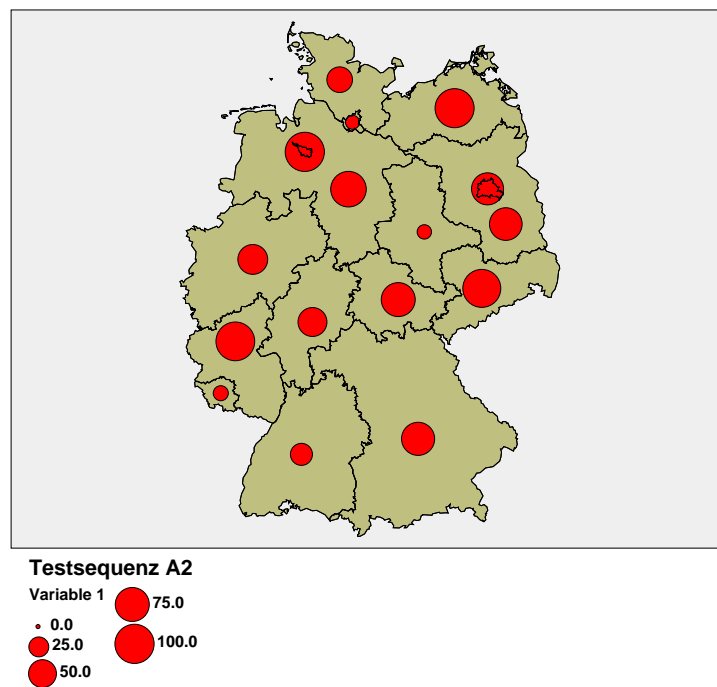
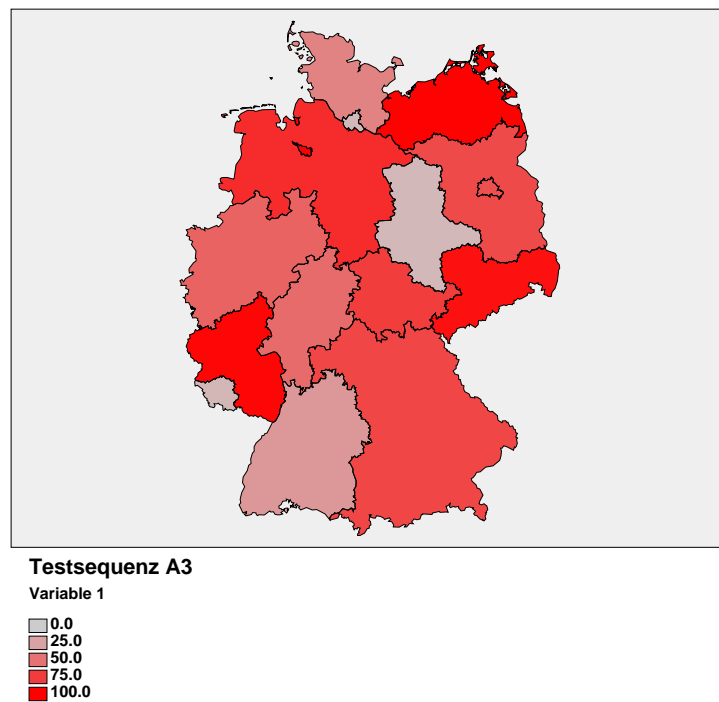
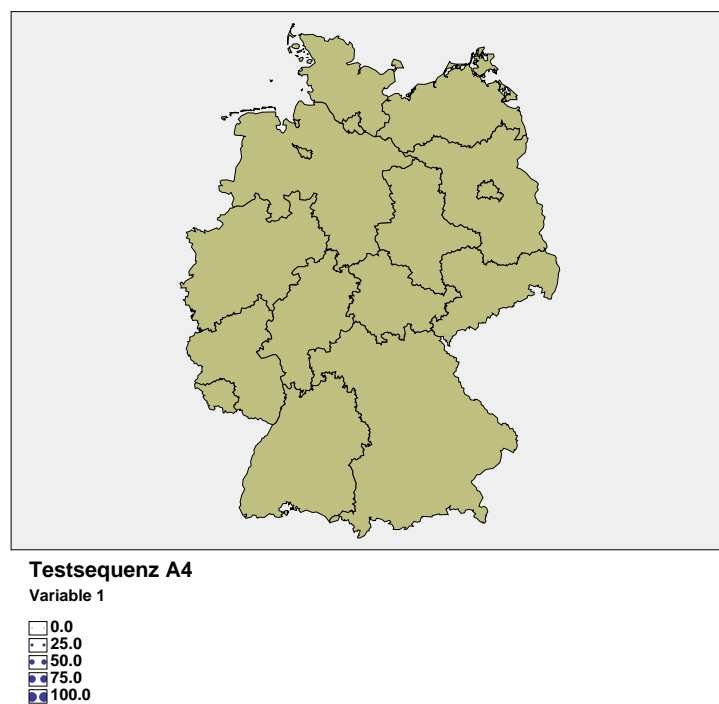


Abb. 6-3: Ein Testbild zur Kreissignaturenkarte



**Abb. 6-4:** Ein Testbild zu Choroplethen/Intensität (im Schwarz-Weiß-Druck ist der Kontrast geringer als in der Originaldarstellung)



**Abb. 6-5:** Ein Testbild zu Choroplethen/Raster

### 6.2.2 Testaufbau zu Werte Finden

Abb. 6-6 zeigt den Aufbau für die Testaufgabe *Werte Finden*. Die Testperson sieht wieder eine thematische Karte der Bundesländer. Die Aufgabe lautet nun, dasjenige Land zu bestimmen, für das ein Wert möglichst nahe bei 50 dargestellt wird. Die Testpersonen werden darüber informiert, daß für je zwei Länder die dargestellten Werte eine Differenz von mindestens 3 haben und daß es immer ein Land mit einem Wert von  $50 \pm 3$  gibt. Damit es bei der interaktiven Auswahl des gefundenen Landes nicht zu Verwechslungen kommen kann, hat die Karte im Fragebogen das gleiche Format und die gleiche Bildschirmposition wie die Karte im Testbild. Die Testpersonen wurden angehalten, während der 5 Sekunden dauernden Darstellungszeit den Cursor auf das gefundene Land zu positionieren und anschließend den Mausknopf zu drücken.

Die Testdaten sind rechteckverteilt in den Grenzen  $[0,100]$  mit der bereits genannten Minimaldistanz. Jede der Sequenzen **F1** - **F4** besteht aus 10 Einzeltests, die die 10 verfügbaren Datensätze in zufälliger Reihenfolge zeigen. Als Fehlermaß dient in den **F**-Tests die Differenz zwischen dem gefundenen Wert und demjenigen, der tatsächlich am nächsten bei 50 liegt.

<p>Testreihe F1, Nr. 1 (Wert finden)</p> <p>Instruktionen:</p> <ul style="list-style-type: none"><li>■ Nachdem Sie auf START geklickt haben, erscheint für einige Sekunden eine Karte.</li><li>■ Versuchen Sie auf dieser Karte dasjenige Land zu finden, bei dem der angezeigte Wert so nahe wie möglich bei 50 liegt.</li><li>■ Anschließend klicken Sie in der Karte rechts auf das Land.</li></ul> <div style="text-align: center; border: 1px solid black; padding: 10px; margin: 10px 0;"><b>START</b></div> <p>Ein paar unverbindliche Versuche gefällig? [Versuch 1] [Versuch 2] [Versuch 3]</p>	<p>Frage:</p> <p>Für welches Land liegt der angezeigte Wert am nächsten zu 50? (Bitte das Land anklicken).</p> <div style="text-align: center;"></div>
--	--

Abb. 6-6: Testaufgabe F: *Wert finden*

### 6.2.3 Testaufbau zu Werte Vergleichen

Abb. 6-7 zeigt den Aufbau für die Testaufgabe **V** (*Werte Vergleichen*). Die Testperson sieht eine thematische Karte der Bundesländer für 5 Sekunden und muß den für Hessen dargestellten Wert mit dem für Mecklenburg-Vorpommern dargestellten Wert vergleichen. Anschließend soll angekreuzt werden, für welches der beiden Länder der höhere Wert gezeigt wurde, falls ein Unterschied erkennbar war.

Die Testdaten sind wieder rechteckverteilt in den Grenzen  $[0,100]$  mit der Einschränkung, daß die Werte für Hessen und Mecklenburg-Vorpommern eine feste Distanz  $d$  haben. Es gibt 10 Testdatensätze, je zwei mit  $d = 5, 4, 3, 2, 1$  (diese Distanzen haben sich in den Vortests als geeignet herausgestellt). Als Fehlermaß dient die Anzahl falscher Antworten pro Sequenz. Bei den Vortests zeigte sich, daß die Aufgabe mit der Visualisierungsform **4** (Punkteraster) praktisch nicht gelöst werden konnte, **V4** wurde deswegen gestrichen.

<p><b>Testreihe VI, Nr. 1</b> (Werte vergleichen)</p> <p><b>Instruktionen:</b></p> <ul style="list-style-type: none"> <li>■ Nachdem Sie auf <b>START</b> gedrückt haben, erscheint für einige Sekunden eine Karte.</li> <li>■ Vergleichen Sie auf dieser Karte die Werte für Hessen und Mecklenburg-Vorpommern und bestimmen Sie das Land mit dem höheren Wert.</li> <li>■ Anschließend beantworten sie die Frage und drücken OK.</li> </ul> <div style="text-align: center; border: 1px solid black; padding: 10px; margin: 20px 0;"> <h1>START</h1> </div> <p>Ein paar unverbindliche Versuche gefällig?  <small>(Versuch 1) (Versuch 2) (Versuch 3)</small></p>	<p><b>Frage:</b> Welches Land hat den höheren Wert für die Variable?</p> <p> <input type="checkbox"/> Mecklenburg-Vorpommern  <input type="checkbox"/> Hessen  <input type="checkbox"/> kein Unterschied erkennbar         </p> <p style="text-align: center;">OK</p>
--	---

**Abb. 6-7:** Testaufgabe V: *Werte Vergleichen*



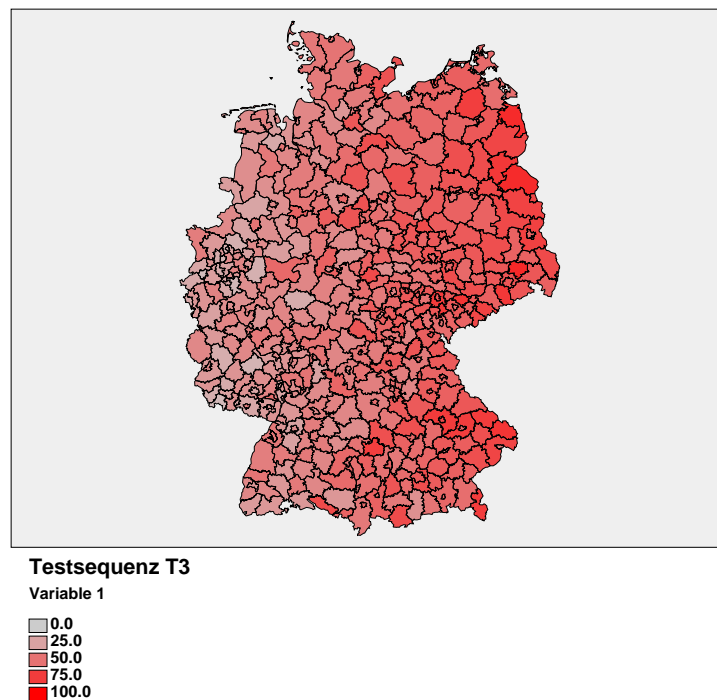
### 6.2.4 Testaufbau zu Trend Erkennen

Abb. 6-8 zeigt den Aufbau für die Testaufgabe *Trend Erkennen*. Wegen der komplexeren Aufgabenstellung wurden die Testbilder hier länger (7 Sekunden) gezeigt. Die Testpersonen waren aufgefordert, in einer thematischen Karte der Landkreise in der Bundesrepublik den geographischen Trend der dargestellten Variable zu erkennen. Sie konnten dazu in einem Menü aus 14 möglichen Verteilungsarten auswählen und waren unterrichtet, daß dem dargestellten Datensatz immer genau eine dieser Verteilungen zugrunde liegt. Abb. 6-9 zeigt beispielhaft ein Versuchsbild aus der Sequenz T3, dem eine Verteilung mit von links nach rechts ansteigenden Werten zugrunde liegt.

Die Testdaten der Sequenzen **T1** - **T4** sind normalverteilt mit  $N(f_i(x, y), \sigma)$ , wobei der Erwartungswert eine Funktion des geographischen Mittelpunktes  $(x, y)$  eines Landkreises ist. Die Funktionen  $f_1$  bis  $f_{14}$  entsprechen den 14 verschiedenen geographischen Verteilungen. Die Testdatensätze wurden erzeugt mit  $\sigma = 24, 32$  und  $40$  (diese Werte für die Standardabweichung haben sich in den Vortests als geeignet herausgestellt). Jedem Testdatensatz liegt also eine der 14 möglichen Verteilungen zugrunde, die je nach der Größe von  $\sigma$  mehr oder weniger stark gestört ist. Als Fehlermaß dient wieder die Anzahl falscher Antworten pro Sequenz.

<p><b>Testreihe T1, Nr. 1</b> (Trend erkennen)</p> <p><b>Instruktionen:</b></p> <ul style="list-style-type: none"> <li>■ Nachdem Sie auf START gedrückt haben, erscheint für einige Sekunden eine Karte.</li> <li>■ Die auf der Karte dargestellte Variable unterliegt einer geographischen Verteilung. Versuchen Sie, den Trend zu erkennen (vergleichen Sie mit den rechts dargestellten Optionen).</li> <li>■ Anschließend beantworten sie die Frage.</li> </ul> <div style="text-align: center; border: 2px solid black; padding: 10px; margin: 10px 0;"> <b>START</b> </div> <p>Ein paar unverbindliche Versuche gefällig?</p> <p>(Versuch 1) (Versuch 2) (Versuch 3) (Versuch 4)  (Versuch 5) (Versuch 6) (Versuch 7) (Versuch 8)  (Versuch 9) (Versuch 10) (Versuch 11) (Versuch 12)  (Versuch 13) (Versuch 14)</p>	<p><b>Frage:</b></p> <p>Welcher geographischen Verteilung unterliegt die Variable?  (Die Werte der Variablen sollen in Pfeilrichtung ansteigen.)</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td><b>WEISS NICHT</b></td> <td></td> </tr> </table>															<b>WEISS NICHT</b>	
		<b>WEISS NICHT</b>															

**Abb. 6-8:** Testaufgabe T: *Trend Erkennen*



**Abb. 6-9:** Versuchsbild aus der Sequenz T3 mit  $\sigma = 8$  (der Trend mit von links nach rechts ansteigenden Werten ist im Schwarz-Weiß-Druck weniger deutlich als in der Originaldarstellung)

### 6.3 Statistische Auswertung

Zweck der statistischen Auswertung ist die Überprüfung der in Kapitel 6.1.1 aufgestellten Hypothesen. Zunächst wird die Hypothese  $H_1$  über die unterschiedliche Effektivität der verschiedenen Visualisierungsformen für jede der Aufgabenstellungen getrennt überprüft. Dazu wird für jede Aufgabenstellung  $A$  die Nullhypothese aufgestellt, daß die Fehlermittelwerte für  $A$  bezüglich der vier Visualisierungsformen identisch sind. Die Alternativhypothese  $H_1^A$  hierzu lautet, daß mindestens zwei der Fehlermittelwerte verschieden sind, d.h. statistisch signifikant voneinander abweichen. Mit der Varianzanalyse wird nun die Nullhypothese überprüft. Falls diese Überprüfung ergibt, daß die Nullhypothese verworfen werden muß, kann die Alternativhypothese angenommen werden.

In einem zweiten Analyseschritt wird nun überprüft, welche der einzelnen Fehlermittelwerte sich von den anderen Fehlermittelwerten signifikant unterscheidet. Dazu dient das LSD-Verfahren (least significant differences) von Fisher. Zur Sicherheit kann diese Überprüfung auch noch mit anderen Verfahren, wie HSD von Tukey oder Newman-Keuls wiederholt werden. Ergebnis dieser Verfahren ist eine Einteilung der Fehlermittelwerte in mehrere homogene Gruppen, wobei die Werte innerhalb einer Gruppe nicht statistisch signifikant voneinander abweichen. Zwischen den Gruppen jedoch können signifikante Unterschiede angenommen werden.

Sind diese zwei Analyseschritte für alle vier Aufgabenstellungen durchgeführt, kann über die eingangs definierten Hypothesen  $H_1$  und  $H_2$  entschieden werden.  $H_1$  ist die Konjunktion der Alternativhypothesen  $H_1^A, \dots, H_1^T$ . Um über  $H_2$  zu entscheiden, müssen die Gruppenstrukturen verglichen werden, die sich aus den LSD-Verfahren zu den vier Aufgaben ergeben. Sind die vier Gruppenstrukturen paarweise verschieden, folgt  $H_2$ .

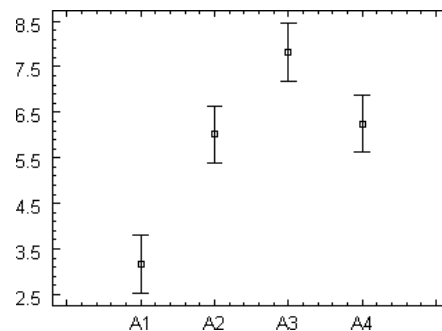
Die Auswertung der Testergebnisse erfolgte mit dem statistischen Analyse- und Visualisierungstool STATLETS [Polh97]. Für die Verfahren wurde ein Konfidenzniveau von 95% festgelegt.

### 6.3.1 Werte Ablesen

Die Varianzanalyse ergibt ein F-Verhältnis von 18.08 und einen p-Wert unter 0.001. Damit weisen die Fehlermittelwerte signifikante Unterschiede auf und  $H_1^A$  kann angenommen werden.

Abb. 6-10 zeigt die Fehlermittelwerte zusammen mit den Intervallen aus dem LSD-Verfahren. Die Intervalle sind derart zu interpretieren, daß bei zwei identischen Mittelwerten deren Intervalle sich in 95% aller Fälle überlappen würden.

Das LSD-Verfahren ergibt drei homogene Gruppen von signifikant unterschiedlichen Mittelwerten. In der ersten Gruppe liegt nur **A1**, in der zweiten Gruppe mit signifikant höheren Werten liegen **A2** und **A4** und in der dritten Gruppe mit nochmals signifikant höheren Werten liegt **A3**. Die Gruppenstruktur ist also: (**{1}**), (**{2, 4}**), (**{3}**).



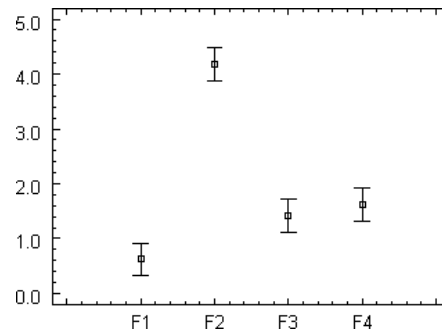
**Abb. 6-10:** Werte Ablesen: Fehlermittelwerte und 95%-LSD-Intervalle

### 6.3.2 Werte Finden

Die Varianzanalyse ergibt ein F-Verhältnis von 50.23 und einen p-Wert unter 0.001. Damit weisen die Fehlermittelwerte signifikante Unterschiede auf und  $H_1^F$  kann angenommen werden.

Abb. 6-11 zeigt die Fehlermittelwerte zusammen mit den 95%-LSD-Intervallen. Das LSD-Verfahren ergibt drei homogene Gruppen von signifikant unterschiedlichen Mittel-

werten. In der ersten Gruppe liegt nur **F1**, in der zweiten Gruppe mit signifikant höheren Werten liegen **F3** und **F4** und in der dritten Gruppe mit nochmals signifikant höheren Werten liegt **A3**. Die Gruppenstruktur ist also: (**{1}**, **{3, 4}**, **{2}**).

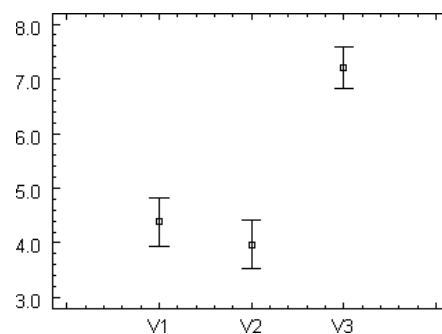


**Abb. 6-11:** Werte Finden: Fehlermittelwerte und 95%-LSD Intervalle

### 6.3.3 Werte Vergleichen

Die Varianzanalyse ergibt ein F-Verhältnis von 39.42 und einen p-Wert unter 0.001. Damit weisen die Fehlermittelwerte signifikante Unterschiede auf und  $H_1^Y$  kann angenommen werden.

Abb. 6-12 zeigt die Fehlermittelwerte zusammen mit den 95%-LSD-Intervallen. Das LSD-Verfahren ergibt zwei homogene Gruppen von signifikant unterschiedlichen Mittelwerten. In der ersten Gruppe liegen nur **V1** und **V2**, in der zweiten Gruppe mit signifikant höheren Werten liegt **V3**. Die Gruppenstruktur ist also: (**{1, 2}**, **{3}**).

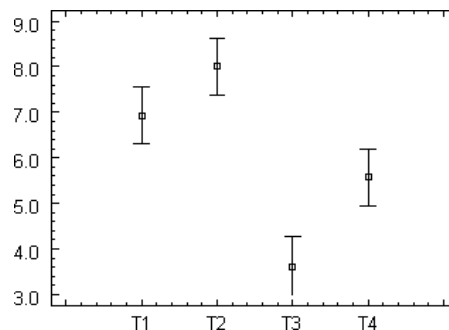


**Abb. 6-12:** Werte Vergleichen: Fehlermittelwerte und 95%-LSD Intervalle

### 6.3.4 Trend Erkennen

Die Varianzanalyse ergibt ein F-Verhältnis von 17.72 und einen p-Wert unter 0.001. Damit weisen die Fehlermittelwerte signifikante Unterschiede auf und  $H_1^T$  kann angenommen werden.

Abb. 6-11 zeigt die Fehlermittelwerte zusammen mit den 95%-LSD-Intervallen. Das LSD-Verfahren ergibt drei homogene Gruppen von signifikant unterschiedlichen Mittelwerten. In der ersten Gruppe liegt nur **T3**, in der zweiten Gruppe mit signifikant höheren Werten liegt nur **T4** und in der dritten Gruppe mit nochmals signifikant höheren Werten liegen **T1** und **T2**. Die Gruppenstruktur ist also: (**{3}**), (**{4}**), (**{1, 2}**).



**Abb. 6-13:** *Trend Erkennen:* Fehlermittelwerte und 95%-LSD Intervalle

## 6.4 Interpretation

### 6.4.1 Werte Ablesen

Im A-Test lautete die Aufgabe, einen Wert möglichst genau aus einer thematischen Karte abzulesen. Als Fehler wurde die Differenz zwischen dem abgelesenen und dem tatsächlich dargestellten Wert definiert. Mit hoher Signifikanz (> 99%) ist die Balkendiagrammkarte die beste Visualisierungsform für diese Aufgabe. Weniger gut geeignet (mit sicherem Abstand) sind die Kreissignaturenkarte und Choroplethen/Raster. Diese beide Visualisierungsformen haben den gleichen Effektivitätsgrad. In nicht sehr großem Abstand darauf folgt Choroplethen/Intensität als schlechteste der getesteten Visualisierungsformen.

Diese Ergebnisse sollen nun in die Sprache der Fuzzy-Effektivität übersetzt werden, damit sie in die Effektivitätsregeln des Entwurfssystems von VIZARD einfließen können. Dabei soll sowohl der ermittelte Effektivitätswert als auch das Konfidenzniveau dieses Wertes einfließen. Für die Übersetzung des Effektivitätsgrades ist eine Verankerung an den Werten good für die beste der getesteten Visualisierungsformen und bad für die schlechteste sinnvoll. Bei den anderen Gruppen sollte die Position relativ zum Gesamtfehlermittel betrachtet werden. Liegt das Mittel innerhalb des LSD-Intervalls der Gruppe kann als Effektivitätswert average angenommen werden.

Für die Übersetzung des Konfidenzniveaus wird beim Standard-Niveau von 95% eine „normal“ sichere Fuzzy-Effektivität zugewiesen. Analog entspricht einem hohen Konfidenzniveau von 99% und damit einem sehr großen Abstand von anderen Fehlermitteln die Qualifizierung *certainly* und einem niedrigen Niveau von 90% die Qualifizierung *probably*.

Zu der Aufgabe *Werte Ablesen* ergibt sich somit für die Balkendiagrammkarte eine Fuzzy-Effektivität von *certainly good*, für Choroplethen/Intensität der Wert *bad*. Die 95%-LSD-Gruppe, bestehend aus Kreissignaturenkarte und Choroplethen/Raster, enthält auch das Gesamtfehlermittel. Diesen beiden Visualisierungsformen wird folglich *average* zugewiesen.

Das gute Ergebnis für die Balkendiagrammkarte entspricht den Erwartungen aus der Theorie: die Länge gilt als effektivste visuelle Variable zur Bestimmung quantitativer Werte. Würde man die Balkendiagramme zusätzlich mit einer Skala ausstatten, was bei mäßiger Signaturdichte möglich ist, könnte das Ergebnis wahrscheinlich sogar noch verbessert werden. Bei den Kreissignaturen (visuelle Variable ist der Flächeninhalt) ist das Ablesen exakter Werte schwieriger. Auch wenn der Flächeninhalt wie in den Testbildern nicht proportional, sondern mit der Quadratwurzel des dargestellten Wertes wächst (wie von Kartographen empfohlen) und mehrere Zwischenwerte in der Legende dargestellt sind, treten deutlich größere Ablesefehler als bei den Balken auf. Das schlechte Ergebnis von Choroplethen/Intensität ist durch verschiedene Faktoren zu erklären. Zum einen können wir deutlich weniger Intensitätsstufen als z.B. Längenstufen unterscheiden, zum anderen kann die Wahrnehmung einer Farbe durch die Farben benachbarter Flächen beeinflusst werden. Bei Choroplethen/Raster spielt womöglich auch die relativ niedrige Auflösung der Bildschirme eine Rolle für die hohen Fehlerraten. Bei der Ausgabe auf Papierkarten dürfte diese Methode effektiver sein, da die wesentlich höhere Auflösung auch feine Unterschiede im Punktraster erkennbar macht.

## 6.4.2 Werte Finden

Im F-Test lautete die Aufgabe, einen festen Wert in einer thematischen Karte zu finden. Als Fehler wurde die Differenz zwischen dem gefundenen und dem in der Aufgabe gesuchten Wert definiert. Mit hoher Signifikanz ( $> 99\%$ ) ist die Kreissignaturenkarte die schlechteste Visualisierungsform für diese Aufgabe. Besser geeignet (mit sicherem Abstand) sind die beiden Formen der Choroplethenkarte. Sie haben den gleichen Effektivitätsgrad. In nicht sehr großem Abstand darauf folgt die Balkendiagrammkarte als bestes der getesteten Visualisierungsformen.

Zu der Aufgabe *Werte Finden* ergibt sich für die Kreissignaturenkarte eine Fuzzy-Effektivität von *certainly bad*, für die Balkendiagrammkarte der Wert *good*. Das LSD-Gruppenintervall von Choroplethen/Intensität und Choroplethen/Raster liegt knapp über dem Gesamtmittel. Diesen beiden Visualisierungsformen wird deshalb *slightly good* zugewiesen.

Das gute Abschneiden der Balkendiagramme und das schlechte Abschneiden der Kreissignaturen ist teilweise darauf zurückzuführen, daß auch zum Finden eines bestimmten

Wertes das exakte Ablesen von Werten in der Karte erforderlich ist. Das Finden von Werten wird aber auch beschleunigt, wenn leicht ein Überblick über die in der Karte dargestellten Werte gewonnen wird und schnell einige Kandidaten für den gesuchten Wert feststehen. Hier ergibt sich ein Vorteil für die beiden Formen der Choroplethen, wie auch das Ergebnis der Aufgabe *Trend Erkennen* zeigt.

### 6.4.3 Werte Vergleichen

Im V-Test lautete die Aufgabe, zwei dargestellte Werte in einer thematischen Karte zu vergleichen und den größeren der beiden Werte zu identifizieren. Als Fehler wurde die Anzahl falscher Antworten definiert. Mit hoher Signifikanz ( $> 99\%$ ) ist Choroplethen/Intensität die schlechtere Visualisierungsform für diese Aufgabe. Besser geeignet (mit sicherem Abstand) sind Balkendiagrammkarte und Kreissignaturenkarte. Sie haben den gleichen Effektivitätsgrad. Schon in den Vortests wurde Choroplethen/Raster abgewertet, da sie zur Lösung der Aufgabenstellung völlig ungeeignet war.

Zu der Aufgabe *Werte Vergleichen* ergibt sich für Choroplethen/Intensität eine Fuzzy-Effektivität von certainly bad, für die Balkendiagrammkarte und die Kreissignaturenkarte der Wert good. Choroplethen/Raster wird mit terrible abgewertet.

### 6.4.4 Trend Erkennen

Im T-Test lautete die Aufgabe, den räumlichen Trend in der Werteverteilung aus einer thematischen Karte abzulesen. Bestes Verfahren ist Choroplethen/Intensität, weniger gut geeignet ist Choroplethen/Raster. Die beiden schlechtesten Verfahren sind Balkendiagrammkarte und Kreissignaturenkarte.

Zu der Aufgabe *Trend Erkennen* ergibt sich somit für Choroplethen/Intensität eine Fuzzy-Effektivität von good, für Balken- und Kreissignaturenkarte der Wert bad. Choroplethen/Raster liegt nahe am Gesamtmittel und erhält average.

Die beiden Formen der Choroplethen sind also den anderen Visualisierungsformen deutlich überlegen, wenn ein guter Überblick über den gesamten Werteverlauf erforderlich ist. Sie eignen sich offensichtlich auch besser zum Erkennen von Mustern und Clustern im Werteverlauf. Dies deckt sich mit der Beobachtung von Bertin, daß für die „höchste“ Stufe der Information die flächenförmigen Signaturen den punkthaften wie Balken und Kreisen überlegen sind.

### 6.4.5 Die zentralen Hypothesen

Die beiden zentralen Hypothesen wurden durch diese Tests bestätigt. Für alle vier Visualisierungsaufgaben wurden signifikante Unterschiede in den Fehlermittelwerten der einzelnen Visualisierungsformen gefunden.  $H_1$  kann deshalb angenommen werden.

Für  $H_2$  müssen die LSD-Gruppenstrukturen der unterschiedlichen Visualisierungsformen betrachtet werden. Ergänzt man **V4** bei *Werte Vergleichen*, ergibt sich folgendes Bild:

*Werte Ablesen:*       ({**1**}, {**2, 4**}, {**3**})  
*Werte Finden:*        ({**1**}, {**3, 4**}, {**2**})  
*Werte Vergleichen:*   ({**1, 2**}, {**3**}, {**4**})  
*Trend Erkennen:*     ({**3**}, {**4**}, {**1, 2**})

Die vier Gruppenstrukturen sind paarweise verschieden.  $H_2$  kann deshalb angenommen werden.

### 6.4.6 Diskussion

Tabelle 6-3 stellt noch einmal alle ermittelten Effektivitätsgrade zusammen, wie sie in den Regeln von VIZARD eingesetzt werden. Abschließend stellt sich die Frage, welchen Gültigkeitsbereich diese Werte haben und inwieweit die Ergebnisse verallgemeinert werden können.

	<i>Werte Ablesen</i>	<i>Werte Finden</i>	<i>Werte Vergleichen</i>	<i>Trend Erkennen</i>
Balkendiagramme	certainly good	good	good	bad
Kreissignaturen	average	certainly bad	good	bad
Choroplethen/Intensität	bad	slightly good	certainly bad	good
Choroplethen/Raster	average	slightly good	terrible	average

**Tabelle 6-3:** Testergebnisse

Diese Werte wurden unter ganz bestimmten Voraussetzungen ermittelt:

- bei einer kleinen, homogen Personengruppe (zwischen 25 und 35 Jahren alt, mit technischer Hochschulausbildung)
- unter Verwendung von Testvisualisierungen, die mit dem Zielsystem (VIZARD) erzeugt wurden
- unter Verwendung von Visualisierungsformen mit kontinuierlichem Signaturverlauf, d.h. ohne Klasseneinteilung des Wertebereichs
- unter Verwendung eines professionellen Farbdisplaygerätes
- unter Gewährung kurzer Zeitspannen für die Bearbeitung

Diese Voraussetzungen müssen beachtet werden, um den Gültigkeitsbereich der Effektivitätsgrade zu bewerten. Wichtig ist zum einen, daß sich die Werte nicht ohne weiteres auf thematische Karten in Atlanten und anderen Printmedien übertragen lassen, die in der Regel mit einem in Klassen aufgeteilten Wertebereich arbeiten. Auch durch die wesentlich höhere Auflösung dieser Medien dürften sich dort etwas andere Effektivitätscharak-



teristika, allerdings mit ähnlicher Tendenz, ergeben. Auch die detaillierte Ausgestaltung der einzelnen Visualisierungsformen, z.B. mit/ohne individueller Werteskala, kann die Effektivität beeinflussen. Die ermittelten Werte lassen sich deswegen uneingeschränkt nur auf das Zielsystem VIZARD übertragen.

Gegen eine weitreichende Verallgemeinerung der Werte in Tabelle 6-3 spricht auch die geringe Zahl und die Zusammensetzung der Testgruppe. Mit den verfügbaren Ressourcen waren keine umfangreicheren Tests möglich. Insbesondere könnte man monieren, daß in der Testgruppe keine ausgesprochenen Laien und keine ausgesprochenen Experten auf dem Gebiet der Analyse raumbezogener Daten vertreten waren. Allerdings sprechen viele Untersuchungen dafür, daß auch zwischen derart unterschiedlichen Gruppen keine signifikanten Unterschiede bei den elementaren visuellen Fertigkeiten und Aufgabenstellungen, wie sie hier getestet wurden, auftreten. Bei Cleveland wurden beispielsweise zwischen einer Gruppe von Hausfrauen ohne technische Vorbildung und einer Gruppe von Personen mit weitreichender Erfahrung in technischen Fragestellungen keine Unterschiede in den elementaren visuellen Fähigkeiten gefunden [ClMc84]. Lewandowsky und Behrens konnten bei der Analyse von Clustern in Karten keine Unterschiede zwischen den Leistungsmerkmalen von Laien und professionellen Statistikern feststellen [LeBe95]. Volbracht untersuchte dreidimensionale Darstellungsmodi anhand von 3D-Modellen chemischer Moleküle und konnte bei den elementaren visuellen Aufgaben keine signifikanten Unterschiede in den Leistungsmerkmalen von Personen mit wenig bzw. viel Erfahrung im Umgang mit Molekülstrukturen feststellen [Volb96].

Schließlich erfordert auch die Festlegung auf eine kurze Bearbeitungszeit von einigen Sekunden eine gewisse Relativierung der Ergebnisse. Durch diese Beschränkung konnten die Tests einfach gehalten und die Vergleichbarkeit der Resultate der einzelnen Personen erhöht werden. Sicher ist eine schnelle Ausführbarkeit der elementaren kognitiven Operationen auch förderlich für die Interpretation des Datensatzes und damit für die Effektivität der Visualisierungsform. Da beim normalen Arbeiten mit GIS diese Zeitbeschränkung aber nicht gegeben ist, sollten die Tests evtl. auch mit anderen Randbedingungen durchgeführt werden. Zwei alternative Formen der Tests sind möglich:

- es wird die Bearbeitungszeit neben dem Bearbeitungsfehler als abhängige Variable betrachtet und mit einer zweifaktoriellen Analyse die Wechselwirkungen dieser Variablen ausgewertet,
- es wird die subjektive Einschätzung der Testpersonen berücksichtigt, d.h. es wird erfragt, welche Visualisierungsformen bei verschiedenen Aufgaben für geeignet gehalten werden.

Als Hauptergebnis dieser Tests kann festgehalten werden, daß zur Lösung einer gegebenen Visualisierungsaufgabe die verschiedenen Visualisierungsformen unterschiedlich gut geeignet sind und daß sich für verschiedene Visualisierungsaufgaben im allgemeinen unterschiedliche Reihenfolgen der Eignung der einzelnen Visualisierungsformen ergeben. Dies stützt einen zentralen Ansatz dieser Arbeit: für den Entwurf effektiver Visualisierungen muß der Kontext, insbesondere die Aufgabenstellung der Visualisierung, beachtet werden. Diese Aussage ist sicherlich auf jede Form von Visualisierung und auf die verschiedensten Aufgabengebiete übertragbar.



# Kapitel 7

## VIZARD: GIS-Visualisierungsumgebung mit Benutzerunterstützung

Kapitel 7 stellt VIZARD, eine prototypische GIS-Visualisierungsumgebung mit integrierter Benutzerunterstützung vor. VIZARD umfaßt Teilrealisierungen aller in dieser Arbeit entwickelten Konzepte: des Modells zur Charakterisierung von Daten und Kontext, der Methodik für den automatisierten Visualisierungsentwurf sowie Ansätze eines integrierten Erklärungs- und Lernsystems. Die Abkürzung VIZARD steht für “Visualization of spatially-related atttribute data”.

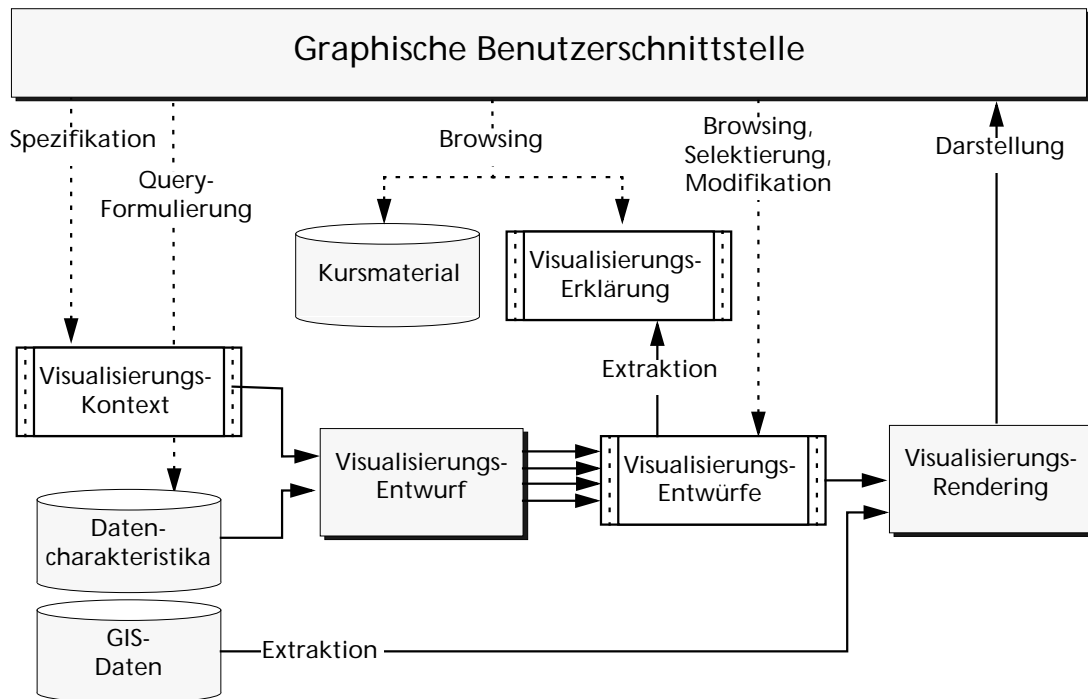
Nach einem Überblick über das Gesamtsystem geht Kapitel 7 auf zwei Aspekte der Realisierung von VIZARD genauer ein: auf die Implementierung der Regelbasis in einer Expertensystemshell und auf die graphische Benutzerschnittstelle der Visualisierungsumgebung. Anhand einer Beispielsitzung wird in die Funktionalität des Systems eingeführt und die Ergebnisse des automatischen Visualisierungsentwurfs im Detail besprochen. Den Abschluß des Kapitels bildet eine Diskussion und Bewertung der Visualisierungsumgebung.

### 7.1 VIZARD

VIZARD wurde mit dem Ziel entwickelt, die Machbarkeit einer intelligenten Visualisierungshilfe für GIS zu zeigen und damit auch die Praktikabilität der zugrundeliegenden Konzepte zu überprüfen. Mit der Untersuchung der Machbarkeit und Praktikabilität sind vor allem die folgenden drei Fragen verbunden:

- Wie gut läßt sich die Entwurfsmethodik für Visualisierungen mit der heute verfügbaren Expertensystemtechnik realisieren?
- Läßt sich eine für interaktive Anwendungen ausreichende Performanz des Entwurfssystems erzielen?
- Wie gut entsprechen die vom System entworfenen Visualisierungen den Spezifikationen des Anwenders?

Die Entwicklung von VIZARD ging vom Historischen Stadtinformationssystem HIST als Basissystem aus, einem spezifischen GIS für die historische Stadtentwicklungsforschung, das schon in Abschnitt 2.1.3 kurz vorgestellt wurde. Den Kern von VIZARD bildet die Expertensystemkomponente für Visualisierungsentwurf, die ausgehend von einer Beschreibung des Kontextes und der Charakteristika des ausgewählten Datensatzes eine Reihe von Visualisierungen entwirft. Anwender sehen von dem System lediglich die graphische Benutzerschnittstelle, die vollständig in HIST eingebettet ist.



**Abb. 7-1:** Systemaufbau und Datenfluß in VIZARD

An Abb. 7-1 lassen sich die typischen Arbeitsschritte mit dem Prototypsystem nachvollziehen. Um einen Datensatz in der Umgebung zu visualisieren, formuliert der Anwender zunächst einen Query auf der Datenbank. Dazu steht ein graphisch-interaktives Werkzeug zur Verfügung. Aus der Datenbank wird dann eine formale Beschreibung des angefragten Datensatzes und seiner Charakteristika extrahiert und an die Entwurfskomponente weitergegeben. Ein weiteres graphisch-interaktives Werkzeug dient zur Spezifikation von Visualisierungszielsetzung und -kontext. Kontextbeschreibung und Zielsetzung werden wie die Datencharakteristika an die Entwurfskomponente als Eingabedaten weitergereicht.

Im allgemeinen entwirft das System für einen Datensatz mehrere ausdrucksfähige Visualisierungen und gibt diese an einen graphischen Browser weiter. Dort kann der Anwender einen Überblick über die verschiedenen Entwürfe erhalten. Sie werden in Reihenfolge ihrer Effektivität präsentiert. Der Anwender kann hier einen passenden Entwurf auswählen und auf das Ausgabemedium, z.B. den Monitor oder einen Drucker rendern lassen. Erst an dieser Stelle wird der Datensatz selbst aus der Datenbank extrahiert, da er für das Ren-

dering benötigt wird. Die Anwender haben auch die Möglichkeit, Entwürfe in einem graphisch-interaktiven Visualisierungseditor zu modifizieren.

Zu einem Visualisierungsentwurf läßt sich jederzeit eine Erklärung anfordern. Sie wird dynamisch aus der Visualisierungsbeschreibung generiert und Hyperlinks zu dem permanent vorhandenen Kursmaterial werden automatisch eingefügt. Für das Studium der Erklärung und des Kursmaterials steht ein Hypertext-Browser zur Verfügung.

## 7.2 Regelbasis

Dieser Abschnitt geht auf die erste Frage der Machbarkeit ein und zeigt, wie sich die wichtigsten Elemente der Entwurfsmethodik für Visualisierungen mit der heute verfügbaren Expertensystemtechnik realisieren lassen. Als Entwicklungsplattform wurde die Expertensystemshell Clips [Giar93] gewählt, die insbesondere im Forschungsbereich weit verbreitet ist. Clips erzeugt Forward Chaining-Systeme, die kompiliert und in C-Programme integriert werden können. Clips selbst hat eine Lisp-orientierte Syntax und erlaubt die Definition von Regeln mit sehr umfangreichem Matchingkriterien.

Für VIZARD wurden 12 der wichtigsten 2D-Visualisierungsformen von GIS berücksichtigt, darunter alle in Kapitel 6 getesteten quantitativen und einige qualitative Kartentypen, sowie mehrere Formen statistischer Graphen. Die prototypische Implementierung umfaßt insgesamt 141 Regeln mit einer durchschnittlichen Länge von 26 Code-Zeilen. Alle Elemente der Entwurfsmethodik konnten als Expertensystemregeln in Clips formuliert werden, mit Ausnahme der Fuzzy-Mengen und ihrer Operatoren, die als Erweiterung von Clips in Form einer C-Bibliothek implementiert wurden. Von allen Regeln der Wissensbasis kann hier nur ein kleiner Teil exemplarisch vorgestellt werden. Die Auswahl fiel daher auf Clips-Realisierungen von Regeln, die schon aus Kapitel 5 bekannt sind.

### 7.2.1 Ausdrucksfähigkeit

Zu jeder Visualisierungstechnik, die das System unterstützt, existieren eine oder mehrere Regeln zur Ausdrucksfähigkeit. Sie implementieren gleichzeitig auch die implizite Dekomposition der Relationen, da sie ein Matching der Datensatzstruktur mit den Ausdruckskriterien durchführen. Die folgende Regel Choropleth::expressiveness illustriert dies am Beispiel der Choroplethen.

Regeln des Expertensystems werden im folgenden immer in einer zweispaltigen Tabelle wiedergegeben. Die linke Spalte enthält die Code-Zeilen der Regel, die rechte Spalte die Erläuterungen. In Clips werden einfache (atomische) Variablen durch ein vorgestelltes Fragezeichen gekennzeichnet, z.B.: ?var, Listenvariablen zusätzlich durch ein vorgestelltes Dollarzeichen, z.B.: \$?liste. Auf der Bedingungsseite der Regeln wird nichtinstantiierten Variablen der matchende Wert zugewiesen. Ist dieser Wert für die Regel ohne Bedeu-

tung, kann auch mit namenlosen Variablen (? bzw. \$?) gearbeitet werden, die dann als Platzhalter fungieren.

(defrule Choropleth::expressiveness:1	Diese Regel überprüft, ob alle Bedingungen der Ausdrucksfähigkeit von Choroplethenkarten auf Teilrelationen des Datensatzes zutreffen. (Vgl. Regel 5-1.)
(dataset (relations \$? ?rel \$?))	<b>1. Bedingung:</b> Der Datensatz enthält mindestens eine Relation.
(relation (name ?rel) (indep-var \$? ?var1 \$?) (dep-var \$? ?var2 \$?) (cardinality 1) )	<b>2. Bedingung:</b> Die Relation enthält mindestens eine unabhängige und eine abhängige Variable und sie hat die Funktionseigenschaft (Kardinalität = 1).
(variable (name ?var1) (type spatial object   reference \$?) (dimension 2) )	<b>3. Bedingung:</b> Die unabhängige Variable ist raumbezogen und es handelt sich dabei um zwei-dimensionale Geo-Objekte oder Geo-Referenzen.
(variable (name ?var2) (type quantitative \$?) (properties \$? ?prop&is-density   is-ratio   is-normalized \$?) )	<b>4. Bedingung:</b> Die abhängige Variable ist quantitativ und bezüglich der Fläche normalisiert oder vom Flächeninhalt unabhängig.
=> (assert (visualization (relation ?rel) (indep-var ?var1) (dep-var ?var2) (name (gensym*)) (type choropleth) (class map) (dimension 2) (why REASON (format nil "%s is spatial & discrete & 2D" ?dom) REASON (format nil "%s is quantitative & %s" ?rng ?prop) "HREF=choropleth.html#Expr")) )	<b>Aktion:</b> Es wird eine neue Visualisierung vom Typ Choroplethenkarte für die Teilrelation var1 → var2 erzeugt.  Im why-Feld wird die vom Erklärungsmodul benötigte Information zur Erläuterung der Ausdrucksfähigkeit dieser Teilvisualisierung abgelegt, sowie ein Hyperlink zum entsprechenden Kapitel des Online-Tutorials.

### 7.2.2 Primitive Effektivität

Die Berechnung der primitiven Effektivität erfolgt gemäß der Methodik in zwei Stufen: zuerst wird die Standardeffektivität ermittelt und diese anschließend modifiziert, falls Er-

eignisse zutreffen. Die Standardeffektivität ergibt sich aus der gewichteten Summe der Einzelbewertungen nach Visualisierungsaufgaben.

Als Beispiel ist hier eine von vier Regeln zur Standardeffektivität von Choroplethen aufgeführt. Sie deckt den Fall der Choroplethen mit Standard-Symbolisierung und bei variablenlokaler Aufgabe ab.

(defrule Choropleth::std-effectiveness:1	Diese Regel bewertet die Standardeffektivität einer Choroplethenkarte mit Standard-Symbolisierung (d.h. kein Punkteraster) und bei variablenlokaler Aufgabe. (Vgl. Regel 5-13.)
(visualization (type choropleth) (name ?viz) (dep-var ?var) (symbolization ~screened-points \$?) ) )	<b>1. Bedingung:</b> Die Visualisierung ist eine Choroplethenkarte mit abhängiger Variable var, die nicht (~) durch ein Punkteraster symbolisiert ist.
(objective (name ?obj) (variables ?var) (task ?task) ) )	<b>2. Bedingung:</b> Die Aufgabe (bzw. eine der Aufgaben) bezüglich der Variable var ist task.
=> (switch ?task	<b>Aktion:</b> Hier wird abhängig von der Aufgabe weiterverfahen:
(case read then (bind ?eff (fuzzy average)) (bind ?why REASON "cannot determine ..." "HREF=choropleth.html#Limits") ) )	Lautet die Aufgabe „Wert ablesen“, wird die Standardbewertung average zugewiesen, zusammen mit den Informationen zur Erläuterung und einem Hyperlink.
(case find then ... ) (case compare then ... ) (case trend then ... ) )	Für die anderen variablenlokalen Aufgaben erfolgt die Zuweisung analog.
(assert (effectiveness (viz ?viz) (type standard) (what ?obj) (eff ?eff) (why ?why))) )	Ein Beschreibung der Standardeffektivität dieser Teilvisualisierung wird aus der zugewiesenen Information erzeugt.

Ist die Standardeffektivität einer Visualisierung bestimmt, erfolgt im nächsten Schritt die Erkennung und Behandlung von Ereignissen. Als Beispiel einer solchen Regel steht hier die Abwertung von Kreissignaturenkarten, die Dichten oder Verhältniszahlen darstellen:

<b>(defrule PropSymbol::event:1</b>	Diese Regel korrigiert die Standardeffektivität der Kreissignaturen nach unten, wenn die dargestellte Variable aus Dichten oder Verhältniszahlen besteht. (Vgl. Regel 5-15.)
(visualization (type prop-symbol) (name ?viz) (dep-var ?var) )	<b>1. Bedingung:</b> Die Visualisierung ist eine Kreissignaturenkarte mit abhängiger Variable var.
(variable (name ?var) (properties \$? is-density   is-ratio \$?) )	<b>2. Bedingung:</b> Die Variable besteht aus Dichten oder Verhältniszahlen.
?fact <- (effectiveness (viz ?viz) (eff ?eff) )	<b>3. Bedingung:</b> Für diese partielle Visualisierung existiert bereits eine Beschreibung der Standardeffektivität.
=> (modify ?fact (effectiveness (viz ?viz) (eff (fuzzy-modify slightly less certain (fuzzy-modify slightly worse ?eff) ) ) ) (why ?why ...))) )	<b>Aktion:</b>  Die Standardeffektivität wird nacheinander mit zwei Fuzzy-Operatoren modifiziert und dadurch eine schlechtere und weniger sichere Bewertung erzielt.

### 7.2.3 Komplexe Effektivität

Als letztes Beispiel für die Realisierung von Regeln der Entwurfsmethodik in einer Expertensystemshell folgt hier noch die Behandlung von hohen Überdeckungsgraden. Ist eine Visualisierung in hohem Grad mit punktförmigen Symbolen bedeckt, wirkt sie unübersichtlich und ihre Effektivität ist gemindert.

<b>(defrule Structure::event:1</b>	Diese Regel wertet Visualisierungen ab, die einen hohen Bedeckungsgrad punktförmiger Symbole haben. (Vgl. Regel 5-19.)
------------------------------------	--



(visualization (type complete) (name ?viz) (symbol-coverage ?cov) )	<b>1. Bedingung:</b> Die Visualisierung ist vollständig und enthält Information über den Bedeckungsgrad.
(test (>= ?cov 0.5))	<b>2. Bedingung:</b> Der Bedeckungsgrad ist größer als 50%.
?fact <- (effectiveness (viz ?viz) (eff ?eff) )	<b>3. Bedingung:</b> Für diese Visualisierung existiert bereits eine Beschreibung der komplexen Effektivität.
=> (modify ?fact(effectiveness (viz ?viz) (eff (fuzzy-modify (shift-ramp ?cov 0.5 1.0 "slightly worse" "much worse") ?eff))) (why ?why ...))) )	<b>Aktion:</b>  Die komplexe Effektivität wird durch einen Fuzzy-Operator modifiziert und je nach Grad der Überdeckung schlechter bewertet. (Zur Definition der Rampenfunktion siehe Seite 141).

## 7.3 Benutzerschnittstelle und Beispielanwendungen

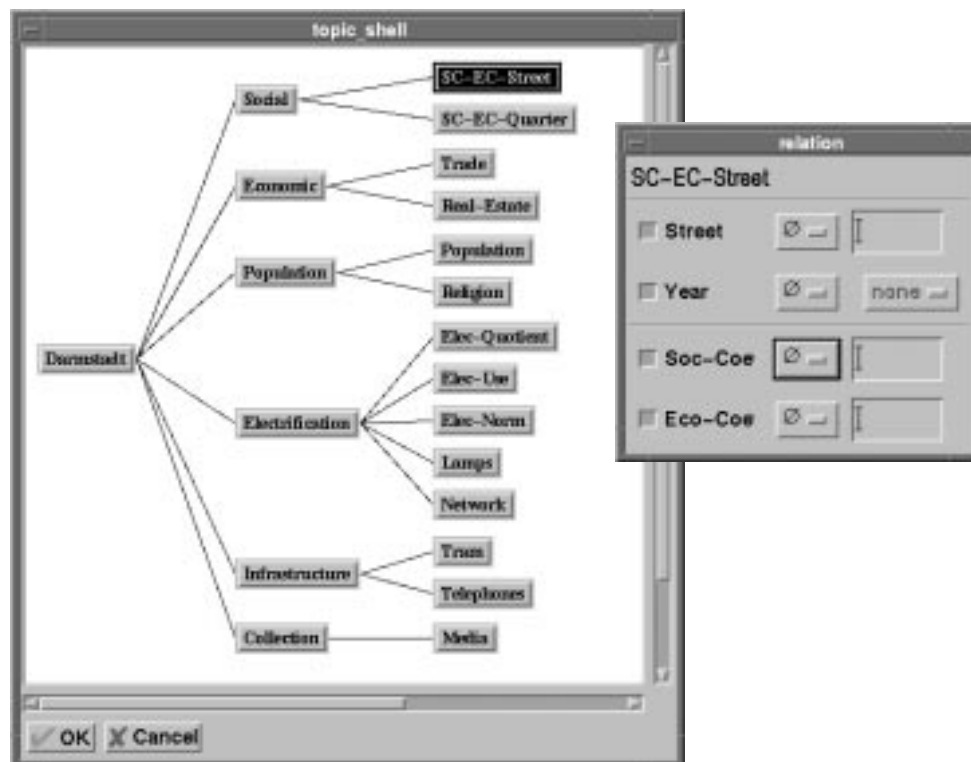
In der Umgebung zur Benutzerunterstützung gibt es mehrere Komponenten, die eine Interaktion mit dem Benutzer erfordern:

- die Formulierung von Geodaten-Queries
- die Spezifikation von Zielsetzung und Kontext
- das Browsing und die Modifikation von Visualisierungen
- das Browsing von Erklärungen und Kursmaterial

Dieses Kapitel stellt die Realisierung der vier Interaktionspunkte anhand von VIZARD vor. Gleichzeitig wird eine beispielhafte Sitzung mit dem System durchgeführt und historische, raumbezogene Daten zur Stadtentwicklung von Darmstadt werden visualisiert. Anhand dieser Beispielsitzung können die Fähigkeiten der Entwurfsmethodik in einem praktischen Einsatzfeld bewertet und ihre Performanz als regelbasiertes System gemessen werden.

### 7.3.1 Spezifikation von Datensatz und Kontext

Soll eine Visualisierung mit VIZARD entworfen werden, spezifiziert der Benutzer zunächst einen Datensatz aus der Datenbank. Dazu steht ein graphischer Browser zur Verfügung, der die gesamte Datenbank in Form eines Baumdiagramms übersichtlich darstellt.



**Abb. 7-2:** Datenbankbrowser und Tabellenbrowser

Die Auswahl von Relationen aus der Datenbasis und von Variablen bzw. Attributen aus einer Relation erfolgt durch Anwählen entsprechender Browser Elemente. Auch komplexere Queries, die Bedingungen an Attributwerte enthalten, werden graphisch spezifiziert.

In der Beispielsitzung wurde ein raumbezogener Datensatz über die Sozial- und Wirtschaftsstruktur der Stadt Darmstadt ausgewählt (siehe Abb. 7-2). Er enthält zwei unabhängige Variablen, *Straße* und *Jahr*, sowie zwei abhängige Variablen, *Sozialkoeffizient* und *Wirtschaftskoeffizient*. Der Sozialkoeffizient ist ein Maß für den sozialen Status der Anwohner einer Straße, er wird über die Berufs- und Statusbezeichnungen in historischen Adreßbüchern ermittelt. Der Wirtschaftskoeffizient ist das Verhältnis der Anzahl von Privatwohnungen zur Anzahl von Geschäften in einer Straße und dient zur Charakterisierung von Wohn- und Geschäftsvierteln einer Stadt. Zur Definition und Verwendung dieser Koeffizienten siehe [BöSc93]. Alle für den Visualisierungsentwurf wichtigen Charakteristika sind in der folgenden Datensatzbeschreibung aufgeführt:

(relation (name SC-EC) (cardinality 1)	Die Relation hat die Funktionseigenschaft (Kardinalität = 1).
--	--

(independent (variable (name Street) (order-type spatial) (data-type reference) (dimension 2) )	Die unabhängige Variable <i>Straße</i> ist geographisch und besteht aus Geo-Referenzen der Dimension 2.
(variable (name Year) (order-type ordinal) (data-type number) (value-range 1880 1910 1940) (properties is-time) )	Die unabhängige Variable <i>Jahr</i> ist ordinal und temporal und sie enthält die drei Werte 1880, 1910, 1940.
(dependent (variable (name SC) (order-type quantitative) (data-type number) (value-domain-bound 0 100) (properties is-ratio) )	Die abhängige Variable <i>Sozialkoeffizient</i> (abgekürzt SC) ist quantitativ und enthält Verhältnis-Zahlen mit den Wertebereichsgrenzen 0 und 100, also Prozentzahlen.
(variable (name EC) (order-type quantitative) (data-type number) (value-domain-bound 0 100) (properties is-ratio) ) ) )	Die abhängige Variable <i>Wirtschaftskoeffizient</i> (abgekürzt EC) ist ebenfalls quantitativ und enthält Prozentzahlen.

Im nächsten Schritt spezifiziert der Benutzer die Aufgaben, die er an die Visualisierung stellt. Dazu steht ihm ein Formular zur Verfügung, in dem er Tasks, Variablen und Domains auswählen sowie Gewichtungen vergeben kann. Die im Beispiel spezifizierte Aufgabenstellung enthält zwei primitive Aufgaben:

(objectives (function exploration (task 60% trend SC) (task 40% compare EC) )	Für SC soll mit einer Gewichtung von 60% der globale Trend gezeigt werden, für EC mit 40% der Wertevergleich ermöglicht werden.
---	---

Auch die Charakteristika des Ausgabemediums lassen sich als weiterer Parameter des Kontextes interaktiv spezifizieren. Im Beispiel (Abb. 7-4) sind die Standardwerte für einen Monitor angegeben.

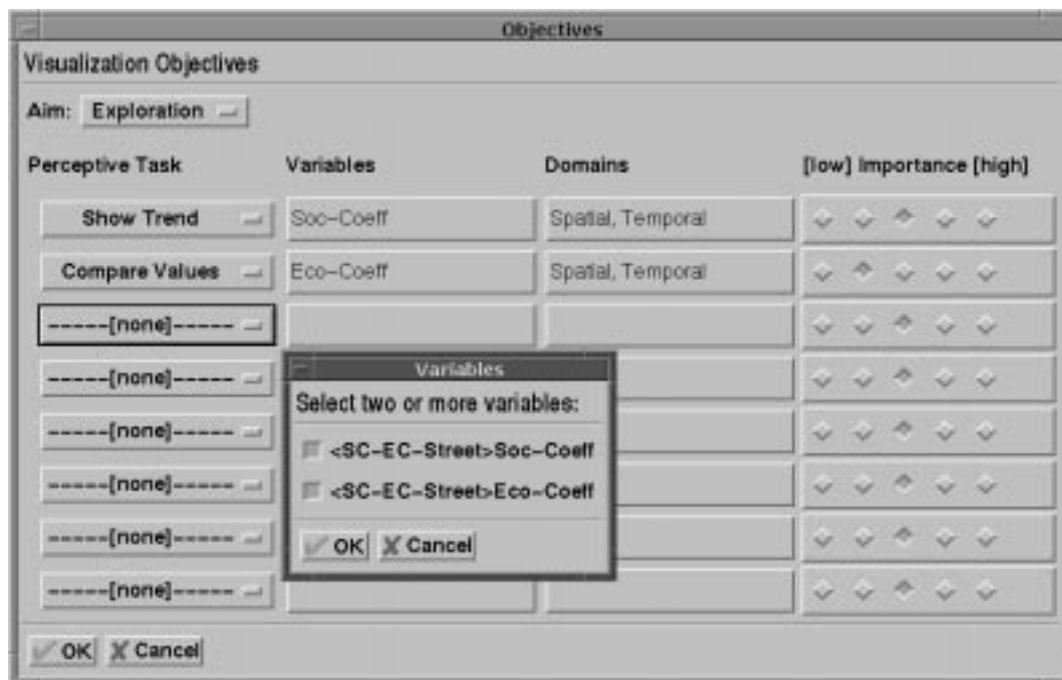


Abb. 7-3: Spezifikation der Aufgabenstellung einer Visualisierung

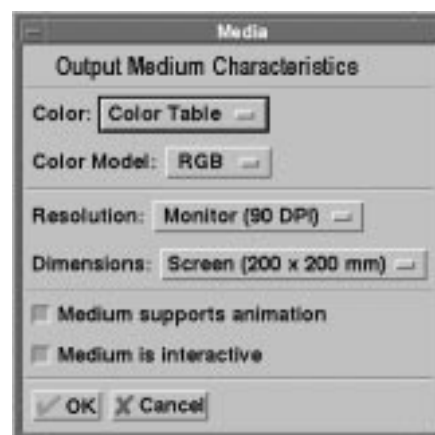
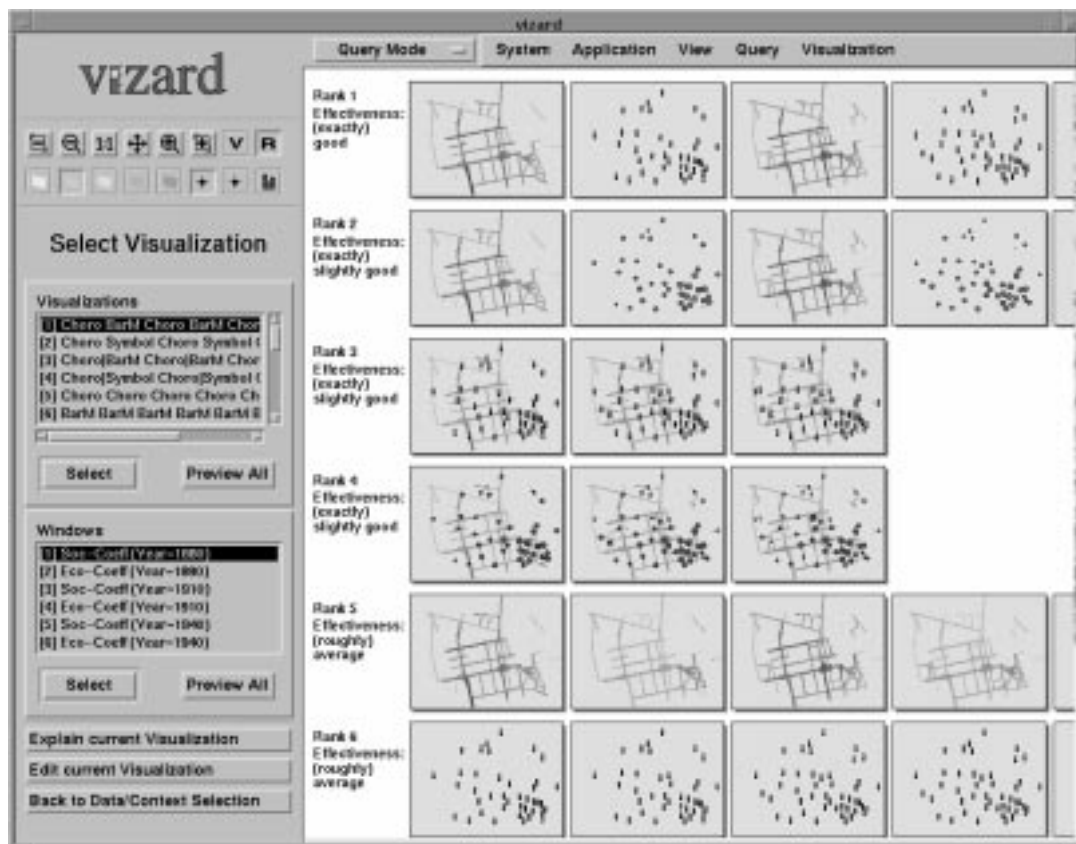


Abb. 7-4: Spezifikation des Ausgabemediums

### 7.3.2 Entwurf und Präsentation von Visualisierungen

Nachdem Datensatz, Zielsetzung und Kontext spezifiziert sind, startet der Benutzer in VIZARD den automatischen Visualisierungsentwurf. In der Regel werden dabei mehrere verschiedene Entwürfe erzeugt, die alle für den Datensatz ausdrucksfähig sind, deren Effektivität bezüglich Zielsetzung und Kontext aber differiert.



**Abb. 7-5:** Präsentation und Auswahl von Visualisierungsentwürfen

In der Beispielsitzung findet VIZARD insgesamt 22 unterschiedliche Visualisierungen des Datensatzes, darunter vier mit der Bewertung *good* oder *slightly good*. Auf einer typischen Workstation (Sun Ultra 1) benötigt das Expertensystem ca. 8 Sekunden, um alle 22 Visualisierungen vollständig zu entwerfen. Dabei feuern insgesamt 1120 Regeln. Für diesen mit vier Variablen schon etwas komplexen Datensatz ist diese Rechenzeit ein sehr guter Wert, der ein interaktives Arbeiten mit dem Expertensystem ermöglicht.

Die sechs besten Visualisierungsentwürfe für den Beispieldatensatz und -kontext werden vom System in einem graphischen Browser angezeigt (siehe Abb. 7-5). Jedes Fenster einer Visualisierung erscheint im Browser stark verkleinert als Icon. Enthält eine Visualisierung mehrere Fenster, werden die entsprechenden Icons nacheinander in einer Zeile angezeigt. Die verschiedenen Visualisierungen erscheinen also als Zeilen einer Matrix in der Reihenfolge ihrer Effektivität, die oberste Zeile von Icons repräsentiert die effektivste Visualisierung. Rechts neben diesen Zeilen sind die entsprechenden Effektivitätsbewertungen in Textform aufgeführt.

Die im Beispiel beste Visualisierung besteht aus sechs Fenstern, je einer Balkendiagrammkarte des Wirtschaftskoeffizienten für die Jahre 1880, 1910 und 1930 und je einer Choroplethenkarte des Sozialkoeffizienten für diese drei Jahre. Wählt man eines der im Browser angebotenen Icons an, stellt das System die entsprechende Visualisierung bzw.

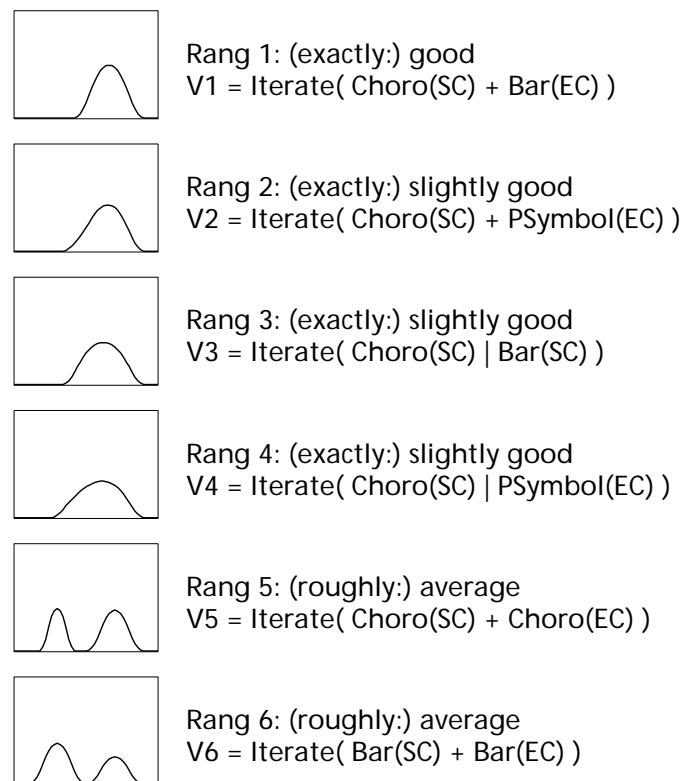


Abb. 7-6: Darstellung und Modifikation von Visualisierungen

ein Fenster der Visualisierung dar. In Abb. 7-6 wurde dasjenige Fenster ausgewählt, das den Sozialkoeffizienten für das Jahr 1910 zeigt.

Warum wurde von VIZARD diese Sequenz von Balkendiagramm- und Choroplethenkarten als beste vollständige Visualisierung für diesen Fall ausgewählt und wie schneiden die anderen Entwürfe in der Bewertung ab? In Abb. 7-7 sind die Fuzzy-Effektivitäten für die sechs besten vollständigen Visualisierungen V1-V6 zum Vergleich untereinander gestellt. Bei allen sechs handelt es sich um Iterationen für die Variable *Jahr*.

Aus der Aufgabenstellung (*Trend Erkennen* für Sozialkoeffizient/SC und *Werte Vergleichen* für Wirtschaftskoeffizient/EC) ergibt sich mit den Standardeffektivitätswerten der Tests, daß SC am besten durch eine Choroplethenkarte (good) und EC am besten durch eine Balken- oder Kreissignaturenkarte (beide ebenfalls good) dargestellt wird. Dies ist bei den vier besten Visualisierungen (V1-V4) der Fall. Da nur variablenlokale Aufgaben anliegen, werden solche Visualisierungen bevorzugt, die jede Variable in einem eigenen Fenster darstellen (V1 und V2). Außerdem wird die Kreissignaturenkarte für EC im Vergleich zur Balkendiagrammkarte leicht abgewertet, da EC aus Verhältniszahlen besteht (Regel 5-15, deswegen V1 vor V2 und V3 vor V4). V5 und V6 liegen wegen der schlechteren primitiven Effektivität dahinter.



**Abb. 7-7:** Fuzzy-Effektivität der sechs besten Entwürfe

Hier noch einmal die Bewertungen im Einzelnen:

### Rang 1

Die primitiven Anteile sind 60% good für die Choroplethen- und 40% good für die Balkendiagrammkarte. Wegen der mehrfachen Visualisierungstechniken (Regel 5-18) ergibt sich eine leichte Abwertung slightly less certain (slightly worse()), hier als worse abgekürzt:

worse (60% good + 40% good)

### Rang 2

Die primitiven Anteile sind 60% good für die Choroplethen- und 40% good für die Kreissignaturenkarte. Die Kreissignaturenkarte wird leicht abgewertet (Regel 5-15). Wegen der mehrfachen Visualisierungstechniken wird insgesamt noch leicht abgewertet:

worse (60% good + worse (40% good))

### Rang 3

Die primitiven Anteile sind 60% good für die Choroplethen- und 40% good für die Balkendiagrammkarte. Überlagerungen werden bei rein variablenlokalen Zielen etwas schlechter bewertet. Wegen der mehrfachen Visualisierungstechniken wird insgesamt

nochmals leicht abgewertet:

worse ( worse (60% good + 40% good))

#### **Rang 4**

Die primitiven Anteile sind 60% good für die Choroplethen- und 40% good für die Kreissignaturenkarte. Die Kreissignaturenkarte wird leicht abgewertet (Regel 5-15). Überlagerungen werden schlechter bewertet und insgesamt wird nochmals abgewertet, da mehrfache Visualisierungstechniken vorliegen:

worse ( worse (60% good + worse (40% good))).

#### **Rang 5**

Hier fließen nur die primitiven Anteile von 60% good für die Choroplethenkarte bei *Trend Erkennen* und 40% certainly bad für die Choroplethenkarte bei *Werte Vergleichen* ein:

60% good + 40% certainly bad

#### **Rang 6**

Hier fließen ebenfalls nur primitive Anteile ein: 60% bad für die Balkendiagrammkarte bei *Trend Erkennen* und 40% good für die Balkendiagrammkarte bei *Werte Vergleichen*:

60% bad + 40% good

#### **Diskussion**

Anhand der Ergebnisse dieser Beispielsitzung lassen sich auch allgemeine Aussagen zur Effektivitätsbewertung der von VIZARD entworfenen Visualisierungen machen. Zunächst kann gesagt werden, daß die Methodik gesicherte Effektivitätsbewertungen liefert, wenn die Standardannahmen zu den Visualisierungsmethoden und zum Kontext zutreffen. In diesem Fall, der bei V1, V5 und V6 zutrifft, werden die experimentell gefundenen Effektivitäten in der entsprechenden Gewichtung verwendet.

Treffen einige der Standardannahmen nicht zu und ist die entsprechende Ereignisbehandlung durch Regeln implementiert, wird eine plausible Effektivitätsbewertung durch teilweise heuristische Verfahren ermöglicht. So wird beispielsweise V2 im Vergleich zu V1 leicht abgewertet, da eine Ereignisregel die Kreissignaturenkarte als nachteilig einschätzt. Durch solche Abwertungen wird vor allem sichergestellt, daß bei ansonsten gleichwertigen Visualisierungen diejenigen vorgezogen werden, auf die keine ungünstigen Ereignisse zutreffen.

Treffen gleichzeitig mehrere Standardannahmen nicht zu und treten dementsprechend viele Ereignisse für eine Visualisierung ein, setzt sich die Effektivitätsbewertung anders zusammen. Die gemessenen Werte sind dann in hohem Maße durch heuristische Regeln angepaßt worden und die Gesamtbewertung ist nicht mehr durch Theorie und Experimente abgesichert. Die Tendenz dieser Gesamtbewertung ist aber auch in diesen Fällen noch aussagekräftig und günstige Visualisierungen werden von ungünstigen unterschieden.



Treten schließlich auch Abweichungen vom Standardkontext auf, die nicht durch Ereignisregeln erfaßt sind, ist die Effektivitätsbewertung nicht mehr aussagekräftig. Damit dieser Fall nicht oder nur sehr selten eintritt, wird man natürlich versuchen, erstens den Standardkontext so zu wählen, daß er möglichst viele Fälle abdeckt und zweitens möglichst alle auftretenden Abweichungen durch Regeln zu erfassen. Dies kann zumindest bei eingeschränkter Aufgabenstellung, also in spezialisierten GIS, erreicht werden. Bei sehr allgemeiner Aufgabenstellung des GIS und stark unterschiedlichen Datencharakteristika und Visualisierungstechniken sollte jedoch mit mehrfachen Standardkontexten gearbeitet werden. Dann kann, z.B. über ein Stereotypenkonzept, der passende Kontext dynamisch ausgewählt werden.

### 7.3.3 Erklärung von Visualisierungen

Erklärungen von Visualisierungsentwürfen und Effektivitätsbewertungen stehen auch den Anwendern von VIZARD zur Verfügung. Das System erzeugt solche Erklärungen bei Bedarf automatisch in Form von Hypertext-Dokumenten im HTML-Standard und zeigt sie mit einem HTML-Browser an (Abb. 7-8 erklärt die Visualisierung V4 aus der Beispielsitzung). Die für die Erklärungen benötigte Information wird schon beim Entwurf und der Auswertung der Visualisierungen generiert, wie an den Regeln in Abschnitt 7.3.2 zu sehen war. Die Erklärungstexte begründen insbesondere, warum eine Visualisierung ausdrucksfähig ist, und wie sich ihre komplexe Effektivität zusammensetzt. Zur Begründung der Ausdrucksfähigkeit werden für jede enthaltene primitive Visualisierung die Ausdruckskriterien zusammen mit den zutreffenden Variablen aufgelistet. Die Ausdruckskriterien erscheinen in dieser Liste nur in Stichworten, um die Erklärung übersichtlich zu halten. Möchte der Anwender detailliertere Information, kann er über ein Hyperlink eine ausführlichere Erklärung anfordern.

Zur Erklärung der Gesamteffektivität einer komplexen Visualisierung werden alle enthaltenen Anteile in schematischer Form aufgelistet. Diese prozentualen Anteile an der Gesamteffektivität werden zusätzlich durch die Anzahl von Balken im Erklärungstext verdeutlicht. Die Farbe der Balken (im Schwarz-Weiß-Druck von Abb. 7-8 nicht zu erkennen) gibt die Tendenz des Anteils wieder: Grün = good, Gelb = average und Rot = bad.

In den gesamten Erklärungstext sind Hyperlinks eingebettet, die auf entsprechende Kapitel des Online-Tutorials über Geodaten-Visualisierung verweisen. Diese Kapitel führen systematisch in die Visualisierungstechniken ein und erläutern ihre Anwendung an Beispielen. Sie enthalten außerdem ausführlichere Beschreibungen der Regeln zur Ausdrucksfähigkeit und Effektivität, die in den Erklärungstexten wie z.B. Abb. 7-8 referenziert werden.

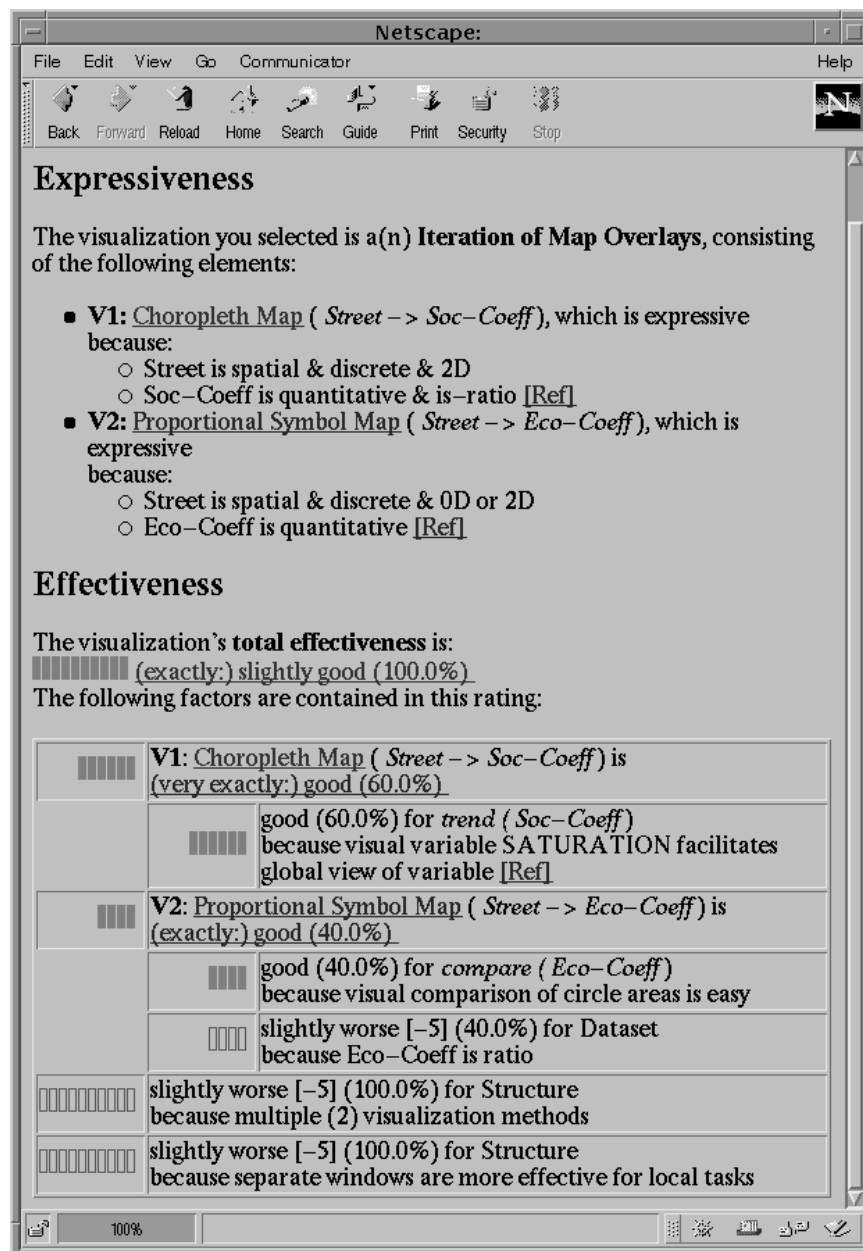


Abb. 7-8: Automatische Erklärung von Visualisierungen

## 7.4 Zusammenfassung und Bewertung

Mit dem Softwaresystem VIZARD wurde die Umgebung zur Benutzerunterstützung in der GIS-Visualisierung prototypisch realisiert. Den Kern des Systems bilden die Regelbasis und das Entwurfssystem, die in der Expertensystemshell Clips implementiert wurden. Mit einem graphischen User-Interface wählt der Benutzer einen Datensatz aus und spezifiziert den Visualisierungskontext. VIZARD entwirft dann automatisch eine Reihe von Visualisierungsvorschlägen. Für den Entwurfsprozeß benötigt das Expertensystem etwa 5 bis 10 Sekunden. Zur Präsentation und interaktiven Auswahl der Visualisierungsentwürfe dient ein graphischer Browser. VIZARD ist in der Lage, jeden Visualisierungsentwurf zu erklären, wozu ein entsprechendes Hypertextdokument automatisch erzeugt und angezeigt wird. Die Erklärungen begründen in Kurzform, warum ein Entwurf ausdrucksfähig ist und welche Faktoren seine Effektivität beeinflussen. Die Erklärungen enthalten außerdem Hyperlinks, die auf ausführlichere Begründungen und grundlegende Gesetze der graphischen Kommunikation verweisen.

Die Realisierung von VIZARD zeigt wesentliche Aspekte der Machbarkeit der Benutzerunterstützungsumgebung:

- Nicht nur die Regeln zur Ausdrucksfähigkeit und Effektivität von Visualisierungen, sondern auch die Entwurfsmethodik selbst (Matching, Komposition, Parameterwahl) kann mit der heute verfügbaren Expertensystemtechnik realisiert werden.
- Die Performanz des wissensbasierten Systems ist ausreichend, um interaktives Arbeiten zu ermöglichen.
- Alle von der Umgebung entworfenen Visualisierungen sind aussagekräftig. Diejenigen Visualisierungen, die vom System als die effektivsten präsentiert werden, entsprechen in der Regel auch den Spezifikationen des Anwenders am besten.

Diese letzte Aussage läßt sich noch weiter präzisieren. Sie bezieht sich auf die Korrektheit der Effektivitätsbewertung in VIZARD. Hier können vier Fälle unterschieden werden:

1. Alle Standardannahmen zu den Visualisierungsmethoden und zum Kontext treffen zu. Dann ist die Effektivitätsbewertung präzise, da sie sich nur auf gemessene oder theoretisch fundierte Werte bezieht.
2. Einige der Standardannahmen treffen nicht zu und die entsprechende Ereignisregeln sind implementiert. Dann wird trotzdem eine gesicherte Effektivitätsbewertung erreicht, da bei ansonsten gleichwertigen Visualisierungen diejenigen vorgezogen werden, auf die keine ungünstigen Ereignisse zutreffen.
3. Es treffen gleichzeitig mehrere Standardannahmen nicht zu und es treten entsprechend viele Ereignisse für eine Visualisierung ein. In diesem Fall ist der exakte Wert der Effektivität zwar nicht mehr gesichert, ihre Tendenz ist aber noch aussagekräftig und günstige Visualisierungen werden richtig erkannt.

4. Treten auch Abweichungen vom Standardkontext auf, die nicht durch Ereignisregeln erfaßt sind, ist die Effektivitätsbewertung nicht mehr aussagekräftig. Dieser Fall kann jedoch durch eine geeignete Modellierung eines oder mehrerer Standardkontexte vermieden werden.

# Kapitel 8

## Zusammenfassung und Ausblick

Das Ziel dieser Forschungsarbeit bestand darin, einen Mangel in der Nutzbarkeit heutiger Geo-Informationssysteme zu beseitigen: sie bieten zwar umfangreiche Funktionalität zur Visualisierung raumbezogener Datensätze, enthalten aber nicht das erforderliche Wissen, um diese Funktionalität sinnvoll zu nutzen. Dieser Mangel äußert sich in der wachsenden Zahl mit GIS erstellter Visualisierungen, die raumbezogene Datensätze ungenau oder sogar verfälscht darstellen und damit räumliche Planung behindern und falsche Entscheidungen hervorrufen können.

Zum Abschluß werden in diesem letzten Kapitel die Ergebnisse der Arbeit zusammengefaßt und ihre Bedeutung für die GIS-Forschung und die Anwendung gezeigt. Anschließend wird ein Ausblick auf zukünftige Arbeiten auf diesem Forschungsgebiet gegeben.

### 8.1 Zusammenfassung der Ergebnisse

Um den als „Wissenslücke in der GIS-Visualisierung“ charakterisierten Mangel zu beheben, wurden Elemente eines integrierten Benutzerunterstützungssystems untersucht, konzipiert und prototypisch realisiert. Der Schwerpunkt dieser Arbeit lag auf der Automatisierung des GIS-Visualisierungsentwurfs unter Einsatz graphischen Wissens und unter Berücksichtigung der Zielsetzung und des Kontexts der Visualisierung.

Vom wissenschaftlichen Standpunkt sind die wichtigsten Ergebnisse dieser Forschungsarbeit:

- ein Modell zur Charakterisierung von Daten und Kontext einer GIS-Visualisierung
- eine wissensbasierte Methodik für den automatisierten, regelbasierten Entwurf von GIS-Visualisierungen
- eine Methodik zur Abschätzung von Visualisierungseffektivität, die sich auf eine Fuzzy-Repräsentation und empirisch bestimmbare Effektivitätscharakteristika stützt

- die prototypische Realisierung eines praktischen Benutzerunterstützungssystems

### Charakterisierungsmodell

Das Charakterisierungsmodell (Kapitel 4) wurde entwickelt, damit die variablen Faktoren der GIS-Visualisierung charakterisiert und in Informationssystemen gespeichert werden können. Es beschreibt die Charakteristika von raumbezogenen Datensätzen und GIS-Kontext und dient als Referenz und wichtiger Ausgangspunkt für die Entwicklung einer Methodik des automatisierten Entwurfs. Das Modell besteht aus Teilen für folgende vier Aspekte:

1. *Datencharakteristika:* Neben den algebraischen und strukturellen Eigenschaften spielt bei GIS-Daten der Raumbezug eine besondere Rolle. Dem wurde durch die Einführung eines geographischen Variablentyps entsprochen, der die klassischen Ordnungstypen (*nominal*, *ordinal* und *quantitativ*) erweitert. Als charakterisierende Eigenschaften wurden außerdem Maßstab, Projektion, geographische Abdeckung und Stetigkeit sowie die Datenqualität herausgearbeitet und formalisiert.
2. *Visualisierungszielsetzung:* Hier lassen sich in der Regel konkrete Aufgabenstellungen der Visualisierung benennen, die durch eine Struktur aus Visualisierungsfunktion, -aufgaben, Variablen und Gewichten formalisiert wurden. Für die GIS-Visualisierung wurden fünf variablenlokale Aufgaben (*Werte Ablesen*, *Finden*, *Vergleichen* und *Trend* bzw. *Cluster Erkennen*) sowie zwei variablenübergreifende Aufgaben (*Variablen Vergleichen* und *Korrelation Erkennen*) vorgeschlagen.
3. *Benutzercharakteristika:* Für die Effektivität einer Visualisierung sind vor allem zwei Eigenschaften der Benutzer bestimmend: visuelle Fähigkeiten wie die Farbwahrnehmung, sowie die Kenntnis Disziplin-spezifischer Visualisierungstechniken.
4. *Ausgabemedien-Charakteristika:* Hierzu zählen Kenngrößen wie Farbtüchtigkeit, Auflösung und Dimensionen, im Falle dynamischer Ausgabemedien auch verschiedene Grade der Interaktivität.

### Entwurfsmethodik

Basierend auf dieses Charakterisierungsmodell wurde eine Methodik für den automatisierten Visualisierungsentwurf in Geo-Informationssystemen entwickelt (Kapitel 5). Ausgehend von der Charakterisierung von Daten und Visualisierungskontext sowie von einem Regelsatz zur Ausdrucksfähigkeit und Effektivität primitiver Visualisierungstechniken werden mit dieser Methodik Visualisierungen eines raumbezogenen Datensatzes in vier Phasen entworfen:

1. *Primitiver Visualisierungsentwurf:* Der Datensatz wird mit den Ausdrucksregeln der Visualisierungstechniken verglichen und es werden entsprechend primitive Visualisierungen von Datensatzteilen erzeugt.
2. *Visualisierungskomposition:* Die primitiven Visualisierungen werden durch Kompositionsoptionen wie Überlagerung und Animation wiederholt zusammengesetzt und alle im Entwurfsraum vorhandenen vollständigen Visualisierungen des Datensatzes gefunden.
3. *Visualisierungsparameter-Optimierung:* Unter Verwendung entsprechender Regeln werden die visuellen Parameter wie Signaturen, Farben und Klassenaufteilungen für jede vollständige Visualisierung optimiert und zugewiesen.
4. *Effektivitätsabschätzung:* Die Effektivität aller vollständigen Visualisierungen wird mit Hilfe von Regeln und einem Auswertungsschema abgeschätzt und in Form von Fuzzy-Mengen modelliert. Die Abschätzung stützt sich auf die gewichtete Summation empirisch bestimmbarer Effektivitätscharakteristika für die verschiedenen Verfahren und Visualisierungsaufgaben in einem Standardkontext.

### Effektivitätstests

Im Rahmen dieser Arbeit wurden auch eine Reihe von Tests zur Bestimmung der Effektivität verschiedener quantitativ-thematischer Kartenformen bei unterschiedlichen Aufgabenstellungen durchgeführt. Ziel dieser empirischen Untersuchung war zum einen die Gewinnung konkreter Werte für die Regeln der Beispielimplementierung VIZARD. Zum anderen sollte mit diesen Tests gezeigt werden, daß auch bei den kartographischen Visualisierungsformen signifikante Unterschiede bezüglich ihrer Eignung für verschiedene Aufgaben bestehen.

Als Hauptergebnis der Tests kann festgehalten werden, daß zur Lösung einer gegebenen Visualisierungsaufgabe die verschiedenen Visualisierungsformen unterschiedlich gut geeignet sind und daß sich für verschiedene Visualisierungsaufgaben im allgemeinen unterschiedliche Reihenfolgen der Eignung der einzelnen Visualisierungsformen ergeben. Dies stützt einen zentralen Ansatz dieser Arbeit: für den Entwurf effektiver Visualisierungen muß der Kontext, insbesondere die Aufgabenstellung der Visualisierung, beachtet werden.

In den Tests konnten für vier quantitativ-thematischer Kartenformen und vier elementare Aufgabenstellungen die relativen Eignungsgrade ermittelt werden, die für das Zielsystem VIZARD Gültigkeit haben. Die Tendenz dieser Effektivitätsgrade läßt sich mit Einschränkung auch auf andere GIS-Umgebungen übertragen.

### VIZARD

Mit dem Softwaresystem VIZARD wurde die Umgebung zur Benutzerunterstützung in der GIS-Visualisierung prototypisch realisiert. Den Kern des Systems bilden die Regelbasis und das Entwurfssystem, die in der Expertensystemshell Clips implementiert wurden. Mit einem graphischen User-Interface wählt der Benutzer einen Datensatz aus und spezi-

fiziert den Visualisierungskontext. VIZARD entwirft dann automatisch eine Reihe von Visualisierungsvorschlägen. Zur Präsentation und interaktiven Auswahl der Visualisierungsentwürfe dient ein graphischer Browser. VIZARD ist in der Lage, einige Aspekte der Visualisierungsentwürfe zu erklären, wozu ein entsprechendes Hypertextdokument automatisch erzeugt und angezeigt wird. Die Erklärungen begründen, warum ein Entwurf ausdrucksfähig ist und welche Faktoren seine Effektivität beeinflussen. Mit der Realisierung von VIZARD konnte gezeigt werden, daß die Benutzerunterstützungsumgebung mit der heute verfügbaren Expertensystemtechnik zu implementieren ist und daß die Performanz für interaktives Arbeiten ausreicht.

## 8.2 Diskussion

### Charakterisierungsmodell

Das Charakterisierungsmodell führt frühere Ansätze zur Charakterisierung von Daten und Kontext für die automatisierte Visualisierung fort. Wichtige Quellen waren [RoMa90] für fundamentale Datencharakteristika, [WeLe90] für Interpretationsziele und [MacE94a] für Geodaten-spezifische Charakteristika.

Als neue und signifikante Aspekte des Modells sind hervorzuheben:

- die aus der Literatur bekannten Ansätze wurden für das Anwendungsgebiet Geodaten-Visualisierung erheblich erweitert
- die verschiedenen Modelle wurden in einem Gesamtmodell konsistent integriert
- basierend auf das Gesamtmodell wurde eine Syntax zur Beschreibung und Speicherung von Daten- und Kontextcharakteristika aufgestellt

Das entwickelte Charakterisierungsmodell kann auch in anderen Bereichen gewinnbringend eingesetzt werden. Viele der beschriebenen charakteristischen Attribute sind nicht nur für den Entwurf von Visualisierungen in GIS, sondern auch in anderen Anwendungsbereichen relevant, so z.B. in der Visualisierung technisch-wissenschaftlicher Datensätze. Andererseits unterstützt insbesondere das Datenmodell aber auch die Konsistenz und Effektivität anderer Operationen in GIS, wie z.B. Analyse- und Verarbeitungsoperationen und die Integration von raumbezogenen Datensätzen aus unterschiedlichen Quellen.

### Entwurfsmethodik

Insbesondere die erste und die zweite Phase der Methodik basieren auf der Arbeit von Mackinlay [Mack86a] und erweitern bzw. adaptieren das von ihm entwickelte Verfahren für die 2D-Visualisierung von Geodaten. Auf dem Gebiet der automatisierten Symbolisierung thematischer Karten wird die Arbeit von Zhang und Buttenfield [ZhBu95] durch die neu entwickelte Methodik so ergänzt, daß die Beschränkung auf univariate Datensätze entfällt und Visualisierungsaufgaben einbezogen werden. Die Fuzzy-Effektivität modifiziert und verallgemeinert die in der Visualisierungshilfe VH [GLLH97] errechneten Eigenschaftsvektoren für Visualisierungsmethoden derart, daß auch Abweichungen vom Standardkontext der Visualisierung berücksichtigt werden können.



Als neue und signifikante Aspekte der Methodik sind hervorzuheben:

- Die Low-Level-Operatoren, die Mackinlay zur Komposition graphischer Primitive beschreibt, werden durch eine Reihe von High-Level-Operatoren zur Komposition von 2D-Visualisierungen von Geodaten ersetzt.
- Visualisierungseffektivität wird erstmals explizit abgeschätzt und repräsentiert. Die Repräsentation erfolgt mittels Fuzzy-Mengen und spezifischer Summations- und Auswertungsoperationen.
- Durch die Fuzzy-Modellierung der Visualisierungseffektivität lassen sich die Zielsetzung und andere Faktoren des Visualisierungskontextes im Visualisierungsentwurf angemessen berücksichtigen und über das Ereignis-Konzept heterogene Effektivitätsregeln integrieren.

Die Verwendung der High-Level-Operatoren zur Visualisierungskomposition ist vorteilhaft für den Entwurf von Geodaten-Visualisierungen, da sie besser geeignet sind, die rigiden Verfahren der thematischen Kartographie zu implementieren. Durch den Wegfall der *Mark Composition* werden unkonventionelle Darstellungen fast vollständig vermieden. Dies bedeutet jedoch auch, daß der Entwurfsraum für Visualisierungen eingeschränkt ist und die insbesondere für das Stadium der explorativen Analyse erwünschte Zahl vielfältiger und kontroverser Sichten des Datensatzes beschränkt wird. Vorteilhaft an den High-Level-Operatoren ist jedoch auch, daß sie sich von der gängigen GIS-Software bereits in gewissem Maße direkt ausführen lassen.

Eine umfassende Regelbasis, als zentralen Bestandteil intelligenter Visualisierungssysteme kann die vorgestellte Methodik nicht ersetzen. In dieser Arbeit wurden Regeln zur GIS-Visualisierung nur für einen kleinen, aber wichtigen Teil der Visualisierungstechniken (die quantitativ-thematischen Karten) und einige exemplarische Aspekte des Visualisierungskontextes aufgestellt. Sie bilden einen Grundstock an Regeln für die intelligente GIS-Visualisierung, müssen jedoch in einem arbeitsintensiven Prozeß der Wissensakquisition ergänzt werden (siehe Abschnitt 8.3.2).

### **Effektivitätsbewertung**

Um eine Aussage über die Korrektheit und Genauigkeit der Effektivitätsbewertung zu treffen, muß zwischen vier Fällen unterschieden werden:

1. Alle Standardannahmen zu den Visualisierungsmethoden und zum Kontext treffen zu. Dann ist die Effektivitätsbewertung präzise, da sie sich nur auf gemessene oder theoretisch fundierte Werte bezieht.
2. Einige der Standardannahmen treffen nicht zu und die entsprechende Ereignisregeln sind implementiert. Dann wird trotzdem eine gesicherte Effektivitätsbewertung erreicht, da bei ansonsten gleichwertigen Visualisierungen diejenigen vorgezogen werden, auf die keine ungünstigen Ereignisse zutreffen.
3. Es treffen gleichzeitig mehrere Standardannahmen nicht zu und es treten entsprechend viele Ereignisse für eine Visualisierung ein. In diesem Fall ist der ex-

akte Wert der Effektivität zwar nicht mehr gesichert, ihre Tendenz ist aber noch aussagekräftig und günstige Visualisierungen werden richtig erkannt.

4. Treten auch Abweichungen vom Standardkontext auf, die nicht durch Ereignisregeln erfaßt sind, ist die Effektivitätsbewertung nicht mehr aussagekräftig.

Damit dieser letzte Fall nicht oder nur sehr selten eintritt, wird man versuchen, den Standardkontext so zu wählen daß er möglichst viele Fälle abdeckt und alle auftretenden Abweichungen durch Regeln zu erfassen. Dies kann zumindest bei eingeschränkter Aufgabenstellung, also in spezialisierten GIS, erreicht werden. Bei sehr allgemeiner Aufgabenstellung des GIS und stark unterschiedlichen Datencharakteristika und Visualisierungstechniken sollte jedoch mit mehreren Standardkontexten gearbeitet werden. Dann kann, z.B. über ein Stereotypenkonzept, der passende Kontext dynamisch ausgewählt werden.

## 8.3 Ausblick

### 8.3.1 Weiterführende Forschungsarbeiten

Zu allen drei Ebenen dieser Arbeit lassen sich Anknüpfungspunkte für weiterführende relevante Forschungsarbeiten identifizieren. Sie wurden teilweise schon an früherer Stelle genannt und sollen hier noch einmal zusammengestellt werden. Zunächst läßt sich das Charakterisierungsmodell an verschiedenen Stellen erweitern und somit für ein breiteres Anwendungsfeld vorbereiten. Ebenso kann man die Methodik erweitern und neue Kompositionsoperationen entwickeln. Schließlich kann auch die prototypische Realisierung der Methodik in Form des Benutzerunterstützungssystems erweitert und durch die Akquisition vielfältiger graphischer Regeln komplettiert werden, damit sie in einem breiten GIS-Kontext Verwendung finden kann.

Betrachtet man das Charakterisierungsmodell, bietet sich als Gegenstand weiterer Forschung insbesondere das Modell der Visualisierungszielsetzung an. Es beschreibt in seiner bisherigen Form zwar einige zentrale Aspekte der Zielsetzung einer Visualisierung, die vor allem in der Phase der explorativen Analyse auftreten können. Es genügt aber nicht mehr zur Beschreibung der komplexen Aufgaben, die Visualisierungen in den späteren Phasen, z.B. beim Testen von Hypothesen und in der Synthese und Präsentation von Ergebnissen unterstützen sollen. Hier könnte z.B. ausgehend von den Ergebnissen Casners [Casn91] in der Aufgabenmodellierung eines eng umgrenzten Anwendungsfeldes, versucht werden, allgemeinere logische Operatoren zu benennen. Ergebnisse solcher Forschungsarbeiten dürften dem automatisierten Visualisierungsentwurf in allen Anwendungsbereichen neue Impulse geben.

Die Methodik zum automatisierten Entwurf von GIS-Visualisierungen läßt sich in mehreren Richtungen weiter ausbauen und vertiefen. Als wichtigste Gebiete weiterer Forschung können identifiziert werden:

- die Entwicklung von Operatoren für die 3D-Komposition und damit die Unterstützung von 3D-Visualisierungstechniken für Geodaten
- die Integration von Verfahren zur kartographischen Generalisierung (als Vorphase der primitiven Visualisierung) und damit die Gewährung einer gewissen Maßstabsunabhängigkeit

VIZARD, die prototypische Realisierung des Benutzerunterstützungssystems läßt sich ebenfalls in mehreren Richtungen weiter ausbauen bzw. für verwandte Anwendungsgebiete adaptieren:

- Adaption und Portierung des Systems für die Unterstützung der Geodaten-Visualisierung im World-Wide-Web.
- Ausbau des graphischen User Interfaces für die integrierte Erklärungs- und Trainingsfunktionalität. Hier dürfte insbesondere die Einführung von UI-Metaphern vielversprechend sein.
- Ausbau der Regelbasis für primitive Visualisierungstechniken, insbesondere für Multiparametertechniken und dreidimensionale Techniken.
- Praktischer Einsatz des Systems zur Unterstützung der Visualisierung in konkreten GIS-Anwendungsprojekten.
- Integration des Systems mit einem kommerziellen Geo-Informationssystem.

### 8.3.2 CommonGIS

CommonGIS (“Common Access to Geographically-Referenced Data”) ist ein Forschungs- und Entwicklungsprojekt, das inhaltlich auf VIZARD und die in dieser Arbeit entwickelte Entwurfsmethodik, sowie auf Arbeiten von Adrienko und Andrienko zur interaktiven Exploration von Geodaten [AnAn97] aufbaut. CommonGIS wird von der Europäischen Kommission (ESPRIT-Programm “Information Access and Interfaces”) vom 1.10.1998 an gefördert. Die Laufzeit beträgt 30 Monate bei einem Volumen von ca. 18 Personenjahren. Neben dem Fraunhofer-Institut für Graphische Datenverarbeitung sind die Gesellschaft für Mathematik und Datenverarbeitung (GMD-FIT), eine GIS-Nutzervereinigung, eine Agentur für Geodaten und zwei Softwarefirmen beteiligt.

Ziel des CommonGIS-Projektes ist, die große Menge statistischer, geo-referenzierter Daten, die von Anbietern weltweit erhältlich sind, einem breiten Nutzerkreis verfügbar zu machen. Dabei sollen zwei wichtige Barrieren für potentielle Nutzer dieser Geodatenbestände überwunden werden: mangelnde Möglichkeiten, auf die Daten zuzugreifen und die Schwierigkeit, die Daten mit den verfügbaren Tools konsistent zu visualisieren und damit eine Interpretation auch für Laien zu ermöglichen.

Dazu werden in diesem Projekt Tools entwickelt, die einen Zugang auf Geodatenbestände über das World-Wide Web ermöglichen und Benutzer bei der Darstellung und Interpretation geo-referenzierter Daten durch ein wissensbasiertes Visualisierungsentwurfssystem

unterstützen. Das Entwurfssystem wird auf die in dieser Arbeit entwickelte Methodik aufbauen. Wichtige Arbeitspakete in CommonGIS sind:

- die systematische Akquisition einer Regelbasis über den Visualisierungsentwurf geo-referenzierter Daten,
- die Entwicklung wissensbasierter Visualisierungsserver für spezielle Geo-Datenbestände in vier europäischen Ländern,
- die Evaluierung des Prototyps durch größere Benutzergruppen und
- die Integration des entstehenden wissensbasierten Tools in ein kommerzielles Geo-Informationssystem und in ein Data Mining-System.

Damit bietet CommonGIS die Möglichkeit, die in der vorliegenden Forschungsarbeit geschaffenen Grundlagen intelligenter GIS-Visualisierungssysteme anzuwenden, unter den verschiedensten Bedingungen zu evaluieren und weiterzuentwickeln. Wenn sich die Konzepte im breiten Einsatz bewähren, wird durch die Integration in kommerzielle Systeme ein neuer Typ von Geo-Informationssystemen geschaffen, die auch Laien die Erzeugung und Interpretation aussagekräftiger Visualisierungen ermöglichen.

# Glossar

**abhängige Variable** eine Variable einer Relation, die von anderen Variablen durch eine funktionale oder mehrwertige Abbildung bestimmt ist; abhängige Variablen ( $A_j$ ) stehen in der Notation von Relationen nach dem Pfeil

**abhängiger Bereich** Produkt der abhängigen Variablen ( $A = (A_1 \times \dots \times A_m)$ ) einer Relation

**ausdrucksfähig (Visualisierung)** Eine Visualisierung ist ausdrucksfähig für einen Datensatz, wenn sie genau die gleiche Information wie der Datensatz enthält

**Ausdrucks-kriterium (einer Visualisierungstechnik)** Bedingung für die Ausdrucksfähigkeit der Visualisierungstechnik

**Benutzermodell** symbolische Beschreibung wesentlicher Eigenschaften eines bestimmten Benutzers oder einer Benutzergruppe

**bivariat (Relation)** mit zwei abhängigen Variablen

**Charakterisierungsschema** systematische Auflistung wesentlicher Eigenschaften eines Gegenstandes

**Choroplethen** Visualisierungstechnik der thematischen Kartographie, bei der ein raumbezogener, quantitativer Datensatz durch eine flächenhafte, gestufte oder stetige Signatur repräsentiert wird

**Datencharakterisierung** symbolische Beschreibung wesentlicher Eigenschaften eines Datensatzes, richtet sich nach dem Charakterisierungsschema für Datensätze

**Datencharakteristika** diejenigen Eigenschaften eines Datensatzes, die für den Visualisierungsentwurf wesentlich sind

**Datenklassifikation** systematische Unterteilung eines quantitativen Wertebereichs in disjunkte Intervalle

**Definitionsbereich-Dekomposition** Dekompositionsoperation, bei der eine unabhängige Variable ersetzt wird

**Defuzzifizierung** Reduktion einer Fuzzy-Menge auf einen scharfen Wert

**Dekomposition (von Relationen)** Aufspaltung einer Relation in Teilrelationen

**effektiv (Visualisierung)** Eine Visualisierung ist effektiv, wenn sie die visuell-kognitiven Fähigkeiten der Zielperson mit den Möglichkeiten des Ausgabemediums optimal ausnutzt

- Effektivität (einer Visualisierung)** eine Bewertung dafür, wie effektiv eine Visualisierung einen Datensatz darstellt
- flächenförmig (Signatur, Objekt)** zweidimensionale, visuelle (bei Signaturen) bzw. symbolische (bei Objekten) Repräsentation räumlicher Gegenstände
- funktionale Dekomposition** Dekompositionsoperation, bei der der unabhängige Bereich aufgespalten wird
- Fuzzy-Effektivität** symbolische Repräsentation der Effektivität einer Visualisierung durch eine Fuzzy-Menge
- Generalisierung** gezielte Veränderung der Information in einer Karte, mit dem Zweck, sie in einem anderen Maßstab darzustellen; eine Stufe des kartographischen Entwurfs
- geographisch (Variable)** ihre Werte sind durch ihre Lage in der Ebene bzw. im Raum geordnet
- Geo-Objekt** symbolische Repräsentation eines Gegenstandes, der eine geographische Lage oder Ausdehnung besitzt
- Geo-Objektdaten** Datensatz, bestehend aus Geo-Objekten
- Geo-Referenz** symbolischer Verweis auf ein Geo-Objekt
- GIS-Visualisierung** 1. eine graphische Darstellung (insbesondere raumbezogener) Information, mit einem GIS erstellt; 2. die Erzeugung graphischer Darstellungen mit einem GIS
- Klassifikation** (siehe *Datenklassifikation*)
- komplex (Visualisierung)** Die Visualisierung enthält mehrere primitive Visualisierungen, die durch Komposition zusammengesetzt wurden
- Komposition (von Visualisierungen)** Operation, die zwei Visualisierungen zu einer neuen, komplexen Visualisierung zusammensetzt
- Kontext** (siehe *Visualisierungskontext*)
- Kreissignaturenkarte** Visualisierungstechnik der thematischen Kartographie, bei der ein raumbezogener, quantitativer Datensatz durch proportional große Kreissignaturen repräsentiert wird
- linienförmig (Signatur, Objekt)** eine zweidimensionale, visuelle (bei Signaturen) bzw. symbolische (bei Objekten) Repräsentation räumlicher Gegenstände
- multivariat (Relation, Datensatz)** mit mehreren abhängigen Variablen
- nominal (Menge, Variable)** ohne Ordnungsrelation
- Objekt** symbolische Repräsentation (insbesondere in einem Computersystem) eines Gegenstandes
- ordinal (Menge, Variable)** linear geordnet, aber ohne Maß
- Ordnungstyp (einer Variable)** Art der Ordnung ihrer Wertemenge: nominal, ordinal, quantitativ oder geographisch

**Parameter** (siehe *Visualisierungsparameter*)

**partiell (Visualisierung)** die Visualisierung drückt nicht alle Variablen eines Datensatzes aus

**primitiv (Visualisierung)** die Visualisierung verwendet eine einzige Visualisierungstechnik

**punktförmig (Signatur, Objekt)** eine eindimensionale, visuelle (bei Signaturen) bzw. symbolische (bei Objekten) Repräsentation räumlicher Gegenstände

**quantitativ (Menge, Variable)** linear geordnet mit präzisen, numerischen Werten

**Query** Anfrage an eine Datenbank, bei der über verschiedene Kriterien Daten ausgewählt werden

**Rasterdaten** Repräsentation eines Ausschnitts der Welt, die auf einer Aufteilung mit einem äquidistanten Gitter beruht

**raumbezogen (Datensatz, Relation)** mindestens eine der enthaltenen Variablen ist geographisch

**raumbezogene Sachdaten** Sachdaten, die Geo-Referenzen enthalten

**Relation** Abbildung  $R : U \rightarrow A$  von Mengen  $U = U_1 \times \dots \times U_n$  (den „unabhängigen Variablen“) in Mengen  $A = A_1 \times \dots \times A_m$  (den „abhängigen Variablen,,)

**Repräsentant (einer Teilrelation)** Softwareobjekt, das alle aus einer Definitionsbereich-Dekomposition entstehenden Teilrelationen repräsentiert

**Sachdaten** symbolische Repräsentation bestimmter Eigenschaften von Gegenständen

**Signatur** genormte, abstrahierte Objektbilder oder Zeichen, die man graphischen Variationen unterziehen kann; Signaturen geben Lage, Qualität und Quantität räumlicher Gegenstände an

**Signatur, gestufte** komplexe Signatur, bei der Quantitätsstufen durch graphische Variation ausgedrückt werden; Beispiel: ein Schema aus Kreissignaturen verschiedener Größe

**Signatur, stetige** komplexe Signatur, bei der eine Quantität durch graphische Variation ausgedrückt wird; Beispiel: eine Kreissignatur mit proportional veränderlicher Größe

**Standardkontext** eine feste Belegung des Kontextes mit typischen Attributen für Benutzer und Ausgabemedium

**statistische Fläche** die Fläche  $\{(x, y, z)\}$ , die entsteht, wenn die Größe eines quantitativen, kontinuierlichen Phänomens an jedem Punkt  $(x, y)$  der geographischen Fläche als Höhe  $z$  interpretiert wird

**Symbolisierung** Zuordnung von Signaturen zu den Attributen eines Datensatzes; eine Stufe des kartographischen Entwurfs

**Technik** (siehe *Visualisierungstechnik*)

**technisch-wissenschaftliche Visualisierung** alle Visualisierungstechniken, die speziell zur Darstellung von Datensätzen aus Simulationen und Messungen in technischen Disziplinen mit Graphikworkstations entwickelt wurden

**thematische Kartographie** Lehre von den kartographischen Techniken zur Darstellung raumbezogener, insbesondere quantitativer, Sachdaten

**unabhängige Variable** eine Variable einer Relation, die nicht von anderen Variablen bestimmt ist; unabhängige Variablen ( $U_i$ ) stehen in der Notation von Relationen nach dem Pfeil

**unabhängiger Bereich** Produkt der unabhängigen Variablen ( $U = (U_1 \times \dots \times U_n)$ ) einer Relation

**univariat (Relation)** mit nur einer abhängigen Variablen

**variablenlokal (Visualisierungsaufgabe)** auf eine einzelne abhängige Variable bezogen

**variablenübergreifend (Visualisierungsaufgabe)** auf mehrere abhängige Variable bezogen

**Vektordaten** Repräsentation eines Ausschnitts der Welt, bei der die Gegenstände durch ihre Position und ihren Rand beschrieben werden

**Visualisierung** 1. Transformation von symbolischer Information in visuelle Information; 2. visuelle Darstellung von Information

**Visualisierungsaufgabe** kognitive Operation oder Interpretationsart, die eine spezifische Visualisierung ermöglichen soll, z.B.: *Werte Vergleichen*, *Cluster Erkennen*

**Visualisierungsentwurf** Auswahl, Komposition und Ausgestaltung von Visualisierungstechniken für Datensätze

**Visualisierungform** (Synonym für *Visualisierungstechnik*)

**Visualisierungskontext** alle variablen Faktoren, außer den Datencharakteristika und der Zielsetzung, die die Effektivität der Visualisierung beeinflussen

**Visualisierungsparameter** alle Parameter einer Visualisierungstechnik, außer dem Datensatz

**Visualisierungstechnik:** eine spezifische Methode zur Visualisierung von Information, parametrisiert durch den Datensatz, den sie abbildet und weitere Parameter, wie die Datenklassifikation und die verwendeten Signaturen

**Visualisierungszielsetzung** die Ziele und Absichten, die mit einer Visualisierung verbunden sind (umfaßt unter anderem die Visualisierungsaufgaben)

**vollständig (Visualisierung)** die Visualisierung drückt alle Variablen des Datensatzes aus

**volumenförmig (Signatur, Objekt)** dreidimensionale, visuelle (bei Signaturen) bzw. symbolische (bei Objekten) Repräsentation räumlicher Gegenstände

**Zielsetzung** (siehe *Visualisierungszielsetzung*)



# Literaturverzeichnis

- [Alle90] R. B. Allen. User models: theory, method and practice. *Int. Journal of Man-Machine Studies*, Vol. 32, 1990, 511–543.
- [AnBa96] L. Anselin, S. Bao. Exploratory Spatial Data Analysis Linking SpaceStat and ArcView. Technical Report 9618, West Virginia University, Morgantown WV, 1996.
- [AnAn98] G. Andrienko, N. Andrienko. Intelligent visualization and dynamic manipulation: two complementary instruments to support data exploration with GIS, *Proceedings Advanced Visual Interfaces'98* (Italy, May 1998), ACM Press, New York, NY, 1998.
- [BaFe81] A. Barr, E. A. Feigenbaum (Eds.). *The Handbook of Artificial Intelligence*, Vol. 1. Kaufmann, Los Altos CA, 1981.
- [Bert74] J. Bertin. *Graphische Semiologie*. de Gruyter, Berlin, 1974. G. Jensch, D. Schade, S. Scharfe. (Übers. und Bearb.).
- [Bert82] J. Bertin. *Graphische Darstellungen und die graphische Weiterverarbeitung von Information*. de Gruyter, Berlin, 1982. W. Scharfe (Übers. und Bearb.).
- [BeFe93] C. G. Beshers, S. K. Feiner. Autovisual: rule-based design of interactive multivariate visualizations. *IEEE Computer Graphics and Applications*, Vol. 13, No. 4, 1993, 41-49.
- [BeFe94] C. G. Beshers, S. K. Feiner. Automated designs of data visualizations. In *Scientific Visualization*, Academic Press, London, 1994, 87–102.
- [BiFr91] R. Bill, D. Fritsch. *Grundlagen der Geo-Informationssysteme*. Band 1: Hardware, Software und Daten. Wichmann, Karlsruhe, 1991.
- [Bill96] R. Bill. *Grundlagen der Geo-Informationssysteme*. Band 2: Analysen, Anwendungen und neue Entwicklungen. Wichmann, Karlsruhe, 1996.
- [BöSc93] H. Böhme, D. Schott. Abschlußbericht für das Projekt „Elektrifizierung und Stadtentwicklung in südwestdeutschen Städten 1880–1935“, Technische Hochschule Darmstadt, Institut für Geschichte, 1993.
- [Brod92] K. W. Brodli et al. *Scientific Visualization: Techniques and Applications*. Springer, Berlin, 1992.

- [BuKu96] B. P. Battenfield, M. P. Kumler. Tools for browsing environmental data: the Alexandria digital library interface. *Proc. 3rd Int. Conf./Workshop on Integrating GIS and Environmental Modeling*. National Center for Geographic Information and Analysis, Santa Barbara CA, 1996.
- [BuMa91a] B. P. Battenfield, D. M. Mark. Expert systems in cartographic design. In D. R. Fraser Taylor (Ed.) *Geographic Information Systems: The Micro-computer and Modern Cartography*. Pergamon Press, Oxford, 1991, 129–150.
- [BuMa91b] B. P. Battenfield, W. A. Mackaness. Visualization. In D. J. Maguire, M. F. Goodchild, D. W. Rind (Eds.). *Geographical Information Systems*, Vol. 1: Principles. Longman, Harlow, Essex, 1991, 427–443.
- [BuSh84] B. Buchanan, E. H. Shortliffe (Eds.). *Rule-Based Expert Systems*. Addison-Wesley, Menlo Park CA, 1984.
- [CaDi87] H. W. Calkins, H. J. Dickson. The effective use of color in cartographic displays. *Proc. of the Int. GIS Symp.*, Vol. 3, Arlington VA, 1987, 189–199.
- [Casn91] S. M. Casner. A task-analytic approach to the automated design of graphical presentations. *ACM Transactions on Graphics*, Vol. 10, No. 2, 1991, 111–151.
- [Cher73] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, Vol. 68, No. 342, 1973, 361–368.
- [Chri91] N. R. Chrisman. The error component in spatial data. In D. J. Maguire, M. F. Goodchild, D. W. Rind (Eds.). *Geographical Information Systems*, Vol. 1. Principles. Longman, Harlow, Essex, 1991, 165–174.
- [CIMc84] W. S. Cleveland, R. McGill. Graphical perception: Theory, experimentation and application in the development of graphical methods. *Journal of the American Statistical Association*, Vol. 79, No. 387, 1994, 531–554.
- [CoHo92] R. E. Cooley, M. H. W. Hoobs. An application of AI to computing class partition values for thematic maps. In *Proc. of the 5th Int. Symp. on Spatial Data Handling*, IGU Commission on GIS, Columbia SC, 1992, 371–380.
- [CoJu98] V. Coors, V. Jung. Using VRML as an interface to the 3D data warehouse. *Proceedings of the VRML'98 Symposium* (Monterey CA, Feb. 16-19, 1998), ACM Press, New York, 1998.
- [Dent96] B. D. Dent. *Cartography: Thematic Map Design*. W. C. Brown Publishers, Dubuque IA, 4th edition, 1996.
- [Denz93] R. Denzer. Graphics for environmental decision-making. *IEEE Computer Graphics & Applications*, Vol. 13, No. 3, March 1993, 58–64.
- [DoGu94] G. O. Domik, B. Gutkauf. User modeling for adaptive visualization systems. *Proc. Visualization '94*, IEEE Computer Society, 1994, 217–223.

- [Domi96] G. O. Domik. Methoden und Techniken der Computervisualisierung, Vortragsreihe an der Technischen Hochschule Darmstadt, 1996. (unveröffentlicht)
- [Dyer90] D. S. Dyer. A dataflow toolkit for visualization. *IEEE Computer Graphics & Applications*, Vol. 10, No. 4, 1990, 60–69.
- [Enca93] J. L. Encarnação et. al. Advanced research and development topics in animation and scientific visualization. In Earnshaw, Watson (Eds.). *Animation and Scientific Visualization: Tools and Applications*, Academic Press, London, 1993.
- [EnSK96] J. L. Encarnação, W. Straßer, R. Klein. *Graphische Datenverarbeitung I*, Oldenbourg Verlag, München, 1996.
- [Ever78] B. S. Everitt. *Graphical Techniques for Multivariate Data*. Heinemann Educational Books, London, 1978.
- [FDFH90] J. Foley, A. van Dam, S. Feiner, J. Hughes. *Computer Graphics: Principle and Practice*, 2nd ed., Addison-Wesley, Reading, MA, 1990.
- [FEDM96] E. D. Fauerbach, R. M. Edsall, D. Barnes, A. M. MacEachren. Visualization of uncertainty in meteorological forecast models. *Proc. 7th Int. Symp. on Spatial Data Handling*, Univ. Delft, Netherlands, 1996.
- [FeMM92] S. Feiner, J. Mackinlay, J. Marks. Automating the Design of Effective Graphics. Tutorial Notes (Tutorial #7 of IEEE Visualization '92, Boston, Oct. 20, 1992)
- [FEER96] G. Fischer, H. Eden, M. Eisenberg, A. Repenning. Domain-Oriented Design Environments: Making Learning a Part of Life. *Communications of the ACM*, Vol 39, No 4, 1996, 40–43.
- [FGDC94] Federal Geographic Data Committee. Content standards for digital geospatial metadata. FGDC, U. S. Geological Survey, Washington, D. C., 1994.
- [FJSD96] L. D. Fosdick, E. R. Jessub, C. J. C. Schauble, G. Domik. *Introduction to High-Performance Scientific Computing*, MIT Press, Cambridge, Mass., 1996.
- [FIHE96] J. Florence, K. Hornsby, M. J. Egenhofer. The GIS WallBoard: interactions with spatial information on large-scale displays. *Proc. 7th Int. Symp. on Spatial Data Handling*, Univ. Delft, Netherlands, 1996.
- [FIRi81] B. Flury, H. Riedwyl. Graphical representation of multivariate data by means of asymmetrical faces. *Journal of the American Statistical Association*, Vol. 76, No. 376, 1981, 757–765.
- [Forr90] D. Forrest. A model of cartographic design for expert system applications. In *Proc. of the 4th Int. Symp. on Spatial Data Handling (Zürich, 1990)*, 753–761, Zürich, 1990. Dept. Geographie, Univ. Zürich.
- [Free91] H. Freeman. Computer name placement. In D. J. Maguire, M. F. Goodchild, D. W. Rind (Eds.). *Geographical Information Systems*, Vol. 1: Principles. Longman, Harlow, Essex, 1991, 445–455.

- [Fran93] A. U. Frank. Geographic information systems: a challenge to computer graphics. State-of-the-Art Report, Eurographics '93, Barcelona, 93.
- [Fran91] A. U. Frank. Design of cartographic databases. In J. C. Muller (Ed.). *Advances in Cartography*. Elsevier Applied Science, London, 1991, 15–45.
- [FrTi94] A. U. Frank, S. Timpf. Multiple representations for cartographic objects in an multi-scale tree - an intelligent graphical zoom. *Computers & Graphics*, Vol. 18, No. 6, 1994, 823–829.
- [Giar93] J. C. Giarratano. CLIPS User's Guide, CLIPS Version 6.0. Technical Report JSC-25013, Johnson Space Center, Houston TX, 1993.
- [GLLS97] K. U. Graw, S. Lange, N. Lopez de Chavez, H. Schumann. Konzept und Realisierung einer intelligenten Visualisierungshilfe. Technischer Bericht CS-08-97, Universität Rostock, Fachbereich Informatik, 1997.
- [GuTD97] B. Gutkauf, S. Thies, G. Domik. A user-adaptive chart editing system based on user modeling and critiquing. *User Modeling: Proceedings of the Sixth International Conference UM97, Sardegna, Italy*, Springer-Verlag, Wien-New York, 1997, 159-170.
- [HaDo96] H. Haase, C. Dohrmann. Doing it right: Psychological tests to ensure the quality of scientific visualization. In M. Göbel, J. David, P. Slavik, J. J. van Wijk (Eds): *Virtual Environments and Scientific Visualization '96*, Springer Verlag, Berlin, 1996, 243-256.
- [Hake82] G. Hake. *Kartographie*. Band I. Sammlung Gröschel, de Gruyter, Berlin.
- [Heal96] C. G. Healey. Choosing effective colours for data visualization. *Proc. IEEE Visualization '96* (held in San Francisco CA, Oct. 27 - Nov. 1, 1996), IEEE Computer Society, 1996.
- [HeJo97] F. Heidmann, M. Johann. Modelling graphic presentation forms to support cognitive operations in screen maps. *Proceedings of the 18th International Cartographic Conference, Stockholm, Sweden*, (Gävle: Swedish Cartographic Society), Vol. 3, 1997, 1452-1461.
- [HeUn94] Hearnshaw, H. M. and Unwin, D. J. (Eds.). *Visualization and GIS*. Wiley, London, 1994.
- [ICA] International Cartographic Association. Commission on visualization, [http://www.gis.psu.edu/ica/icavis/ICAvs\\_overview.html](http://www.gis.psu.edu/ica/icavis/ICAvs_overview.html), (World-Wide Web, ohne Datum).
- [IEEE92] IEEE Visualization '92, Workshop on Automated Design of Visualizations, Workshop Notes, 1992. (unveröffentlicht)
- [Imho71] E. Imhof. *Thematische Kartographie*. de Gruyter, Berlin, 1971.
- [Imho75] E. Imhof. Positioning names on maps. *The American Cartographer*, Vol. 2, No. 2, 1975, 128–144.

- [Jaco94] B. Jacobs. Der Einfluß von Graphtyp und Graphanordnung auf das Graphverstehen bei der Analyse von Verläufen. Arbeitsbericht Nr. 13, Medienzentrum der Philosophischen Fakultät, Universität des Saarlands, Saarbrücken, 1994.
- [JGLT95] V. Jung, C. Giger-Hofmann, K. Lutze, and J. C. Teixeira. Efficient Spatial Data Acquisition in Temporal Geographic Information Systems. In J. C. Teixeira and J. Rix (Eds.), *Modeling and Graphics in Science and Technology*. Springer, Berlin, 1995.
- [JiBO96] B. Jiang, A. Brown, F.-J. Ormeling. Some perceptual aspects of colouring uncertainty. *Proc. 7th Int. Symp. on Spatial Data Handling*, Univ. Delft, Netherlands, 1996.
- [Jone89] C. B. Jones. Cartographic name placement with Prolog. *IEEE Computer Graphics & Applications*, Vol. 9, No. 9, 1989, 36–47.
- [Jung93] V. Jung. Visualization of flow simulation data in environmental modeling. In M. Göbel and J. C. Teixeira (Eds.), *Graphics Modeling and Visualization in Science and Technology*. Springer, Berlin, 1993, 74–84.
- [Jung94] V. Jung. A geographic information system for the support of historic research. *Computers & Graphics*, Vol. 18, No. 2, 1994, 219–225.
- [Jung95] V. Jung. Knowledge-based visualization design for geographic information systems. In *Proc. of the 3rd ACM Int. Workshop on Advances in Geographic Information Systems (Baltimore MD, 1995)*, ACM Press, New York, 1995, 101–108.
- [Jung96a] V. Jung. Guiding users in the visualization of geographic data. In *1st Int. Conf. on GeoComputation*, Leeds, 1996, 470–482.
- [Jung96b] V. Jung. Fuzzy effectiveness evaluation for intelligent user interfaces to GIS visualization. In *Proc. of the 4th ACM Int. Workshop on Advances in Geographic Information Systems (Rockville MD, 1996)*, ACM Press, New York, 1996, 157–164.
- [KeKe92] P. R. Keller, M. M. Keller. *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press, Los Alamitos CA, 1992.
- [KoSi86] H. F. Korth and A. Silberschatz. *Database System Concepts*. McGraw-Hill, New York, 1986.
- [KoVo91] V. J. Kollias, A. Voliotis. Fuzzy reasoning in the development of geographical information systems - FRIS: a prototype soil information system with fuzzy retrieval capabilities. *Int. Journal of Geographical Information Systems*, Vol. 5, No. 2, 1991, 209–223.
- [Kraa90] M.-J. Kraak. Theoretical aspects of three-dimensional cartography. *Int. Yearbook of Cartography*, Vol. XXX, 1990, 81–91.
- [Kres94] W. Kresse. *Plazierung von Schrift in Karten*. Dissertation (Dr.-Ing.), Landwirtschaftliche Fakultät, Universität Bonn, 1994.

- [KrMa94] M.-J. Kraak, A. M. MacEachren. Visualization of the temporal component of spatial data. *Proc. of the 6th Int. Symp. on Spatial Data Handling*, Edinburgh, 1994, 391–409.
- [KuFr90] W. Kuhn, A. U. Frank. A formalization of metaphors and image-schemas in user interfaces. In D. M. Mark, A. U. Frank (Eds.), *Cognitive and Linguistic Aspects of Geographic Space*, Kluwer Academic, Dordrecht, 1990, 419–434.
- [Lane93] D. M. Lane. *HyperStat: Hypermedia for Learning Statistics and Analyzing Data*, Academic Press, New York, NY, 1993.
- [Lang92] G. E. Langram. *Time in Geographic Information Systems*. Taylor & Francis, London, 1992.
- [LSMK95] S. Lange, H. Schumann, W. Müller, D. Krömker. Problem-oriented visualization of multi-dimensional data sets. *Proc. ISSV '95*, Chia, Italy, 1995.
- [LeBe95] S. Lewandowsky, J. T. Behrens. Visual detection of clusters in statistical maps. *Proceedings of the 1995 Meeting of the American Statistical Association*, Orlando, Florida, 1995.
- [LBMP96] C. S. Li, D. Bree, A. Moss, J. Petch. Developing internet-based user interfaces for improving spatial data access and usability. *Proc. 3rd Int. Conf./Workshop on Integrating GIS and Environmental Modeling*. National Center for Geographic Information and Analysis, Santa Barbara CA, 1996.
- [LiSa92] M. Lindholm, T. Sarjakowsky. User models and information theory in the design of a query interface for GIS. In A. U. Frank, I. Campari, U. Formentini (Eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Springer, Berlin, 1992, 328–347.
- [MacE94a] A. M. MacEachren. *Some Truth with Maps: A Primer on Symbolization & Design*. Association of American Geographers, Washington DC, 1994.
- [MacE94b] A. M. MacEachren. Time as a cartographic variable. In H. M. Hearnshaw, D. J. Unwin (Eds.) *Visualization and GIS*, Wiley, London, 1994, 115–130.
- [MaTa94] A. M. MacEachren, D. R. F. Taylor (Eds.). *Visualization in Modern Cartography*, Pergamon, Oxford, 1994.
- [Mack86a] J. Mackinlay. *Automatic Design of Graphical Presentations*. PhD thesis, Stanford University, Computer Science Dept, Stanford CA, 1986.
- [Mack86b] J. Mackinlay. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. Graph.*, Vol. 5, No. 2, April 1986, 110–141.
- [Mack91] J. Mackinlay. Search Architectures for the Automatic Design of Graphical Presentations. In J. W. Sullivan and W. T. Sherman (Eds.), *Intelligent User Interfaces*, ACM Press, New York, 1991, 281–292.
- [MaGR91] D. J. Maguire, M. F. Goodchild, D. W. Rind (Eds.). *Geographical Information Systems*, Vol. 1: Principles. Longman, Harlow, Essex, 1991.

- [Medy94] D. Medyckyj-Scott. Visualization and human-computer interaction in GIS. In H. M. Hearnshaw and D. J. Unwin (Eds.), *Visualization and GIS*, Wiley, London, 1994, 200–211.
- [MeGr88] G. W. Meyer, D. P. Greenberg. Color-Defective Vision. *IEEE Computer Graphics & Applications*, Vol. 9, September 1988, 28-40.
- [Monm90] M. Monmonier. Strategies for the interactive exploration of geographic correlation. *Proc. of the 4th Int. Symp. on Spatial Data Handling* (Zürich, 1990), Dept. Geographie, Univ. Zürich. Zürich, 1990, 512–521.
- [Monm91] M. Monmonier. *How to Lie with Maps*. Univ. of Chicago Press, Chicago IL, 1991.
- [Monm92] M. Monmonier. Authoring graphic scripts: experiences and principles. *Cartography and Geographic Information Systems*, Vol. 19, No. 4, 1992, 247–262.
- [MuJV86] J.-C. Muller, R. D. Johnson, and L. R. Vanzella. A knowledge-based approach for developing cartographic expertise. In *Proc. of the 2nd Int. Symp. on Spatial Data Handling* (Seattle WA, 1986), IGU Commission on GIS, Williamsville NY, 1986, 557–571.
- [Mull93] J.-C. Muller. Latest developments in GIS/LIS. *Int. Journal of Geographical Information Systems*, Vol. 7, No. 4, 1993, 293–303.
- [Müll98] W. Müller. *Effektiver Einsatz grundlegender Darstellungsprimitive zur Informationsvisualisierung*. Dissertation (Dr.-Ing.), Fachbereich Informatik, Technische Universität Darmstadt, 1998. (im Druck)
- [Neve88] R. J. J. Neves. Flow process modeling in a salt marsh. *Proc. Computer Modeling in Ocean Engineering*, 1988, 303–309.
- [OlBr97] J. Olson, C. Brewer. An evaluation of color selections to accommodate map users with color-vision impairments, *Annals of the Association of American Geographers*, Vol. 87, No. 1, 1997.
- [Polh97] N. W. Polhemus. STATLETS User Manual, NWP Associates, Princeton, N. J., 1997, <http://www.statlets.com/usermanual>.
- [RhLa95] P. Rheingans, C. Landreth. Perceptual principles for effective visualizations. In G. Grinstein, H. Levkowitz, (Eds.), *Perceptual Issues in Visualization*, Springer-Verlag, Berlin, 1995, 59-74.
- [Robe91] P. K. Robertson. A methodology for choosing data representations. *IEEE Computer Graphics & Applications*, Vol. 11, No. 3, March 1991, 56–67.
- [RoCM93] G. G. Robertson, S. K. Card, J. D. Mackinlay. Information visualization using 3D interactive animation. *Comm. ACM*, Vol. 36, No. 4, 1993.
- [RoFe94] P. K. Robertson, L. De Ferrari. Systematic approaches to visualization: is a reference model needed? In *Scientific Visualization*, Academic Press, London, 1994, 287–305.

- [RoMa90] S. F. Roth and J. Mattis. Data characterization for intelligent graphics presentation. In *Proc. of CHI'90 (Seattle WA, 1990)*, ACM Press, New York, 1990, 193–200.
- [ScJu94] D. Schott, V. Jung. HIST: Ein Geo-Informationssystem für Stadtgeschichte. In *Erste Arbeitstagung Geschichte und EDV*, AG Geschichte und EDV, Göttingen, 1994.
- [ScLG96] H. Schumann, N. López de Chávez, K.-U. Graw. Visual representation of multiparameter data with spatial dependence. *Proc. 7th Eurographics Workshop on Visualization in Scientific Computing*, Prague, 1996.
- [ScUr97] H. Schumann, B. Urban. Evaluation of marine data by visual means. *Proc. 8th Eurographics Workshop on Visualization in Scientific Computing*, Boulogne sur Mer, France, 1997.
- [SDKR95] W. Smith, J. Dunn, K. Kirsner, M. Randell. Colour in map displays: issues for task-specific design. *Interacting with Computers*, Vol. 7, No. 2, 1995, 151–165.
- [SeFe91] D. D. Seligman, S. Feiner. Automated generation of intent-based 3D illustrations, *Computer Graphics*, Vol. 25, No. 4, July 1991, 123–132.
- [SeIg90a] H. Senay and E. Ignatius. VISTA: Visualization tool assistant for viewing scientific data. In *SIGGRAPH'90 Course Notes, No. 27 (State of the Art in Data Visualization)*, 1990.
- [SeIg90b] H. Senay and E. Ignatius. Rules and principles of scientific data visualization. Technical Report No. GWU-IIST-90-13, Dept. of Electrical Eng. and Computer Science, The George Washington Univ., Washington, D.C., 1990. Auch in *State of the Art in Data Visualization*, Tutorial Notes, Siggraph '90.
- [SeIg91] H. Senay and E. Ignatius. Compositional analysis and synthesis of scientific data visualization techniques. In N. M. Patrikalakis (Ed.), *Scientific Visualization of Physical Phenomena*, 269–281. Springer, Berlin, 1991.
- [SeIg94] H. Senay and E. Ignatius. A knowledge-based system for visualization design. *IEEE Computer Graphics & Applications*, Vol. 14, No. 6, Nov. 1994, 36–47.
- [SlEg91] T. A. Slocum, S. L. Egbert. Cartographic data display. In D. R. Fraser Taylor (Ed.) *Geographic Information Systems: The Microcomputer and Modern Cartography*. Pergamon Press, Oxford, 1991, 167–199.
- [SyMC96] J. Symanzik, J. J. Majure, D. Cook. The linked ArcView 2.1 and XGobi environment: GIS, dynamical statistical graphics, and spatial data. In *Proc. of the 4th ACM Int. Workshop on Advances in Geographic Information Systems (Rockville MD, 1996)*, ACM Press, New York, 1996, 155–162.
- [Till92] T. Tilli. *Fuzzy-Logik: Grundlagen, Anwendungen, Hard- und Software*. Francis-Verlag, München, 1992.



- [Tuft83] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire CT, 1983.
- [Tuft90] E. R. Tufte. *Envisioning Information*. Graphics Press, Cheshire CT, 1990.
- [Turk90] A. Turk. Towards an understanding of human-computer interaction aspects of geographic information systems. *Cartography*, Vol. 19, No. 1 (June), 1990, 31–60.
- [Turk94] A. Turk. Cogent GIS Visualizations. In H. M. Hearnshaw and D. J. Unwin (Eds.), *Visualization and GIS*, Wiley, London, 1994, 26–33.
- [Tyne92] J. Tyner. *Introduction to Thematic Cartography*. Prentice Hall, Englewood Cliffs NJ, 1992.
- [Volb96] S. Volbracht. Empirische Evaluierung dreidimensionaler Darstellungsmodi, Workshop Bits -> Bilder -> Bedeutung. In *Geoinformatik online*, <http://gio.uni-muenster.de>, Dezember 1996.
- [WaHS90] F. Wang, G. B. Hall, Subaryono. Fuzzy information representation and processing in conventional GIS software: database design and application. *Int. Journal of Geographical Information Systems*, Vol. 4, No. 3, 1990, 261–283.
- [Wang92] Z. Wang. An Intelligent Interface for Application of Graphic Elements. In *Proc. of the 5th Int. Symp. on Spatial Data Handling*, IGU Commission on GIS, Columbia SC, 1992, 391–400.
- [Ware88] C. Ware. Color sequences for univariate maps: theory, experiments, and principles. *IEEE Computer Graphics & Applications*, Vol. 9, September 1988, 41–49.
- [WaWo95] F. Wagner, A. Wolff. Map labeling heuristics: provably good and practically useful. *Proc. of the 11th Annual ACM Symp. on Computational Geometry*, ACM Press, New York, 1995, 109–118.
- [Weib91] R. Weibel. Amplified Intelligence and Rule-Based Systems. In B. P. Buttenfield and R. B. McMaster (Eds.), *Map Generalization: Making Rules for Knowledge Representation*, 172–186. Longman, London, 1991.
- [Weib94] R. Weibel. Geographic information systems (GIS) and visualization. State-of-the-Art Report, Eurographics '94, Oslo, 1994.
- [WeBu92] R. Weibel, B. P. Buttenfield. Improvement of GIS Graphics for Analysis and Decision-making. *Int. Journal of Geographical Information Systems*, Vol. 6, No. 2, 1992, 223–245.
- [WeGG96] F. J. M. van der Wel, L. C. van der Grag, B. G. H. Gorte. Visual exploration of uncertainty in remote sensing classifications. In *1st Int. Conf. on GeoComputation*, Leeds, 1996, 843–856.
- [WeLe90] S. Wehrend, C. Lewis. A problem-oriented classification of visualization techniques. *Proc. Visualization '90*, IEEE Computer Society Press, 1990.

- [Witt91] R. Wittenberg. *Grundlagen computergestützter Datenanalyse*. Handbuch für computergestützte Datenanalyse, Band 1, R. Wittenberg, H. Küffner (Hrsg.), Gustav Fischer Verlag, Stuttgart, 1991.
- [WoBW96] J. D. Wood, K. W. Brodlie, H. Wrigh. Visualization over the World Wide Web and its application to environmental data. *Proc. IEEE Visualization '96* (held in San Francisco CA, Oct. 27 - Nov. 1, 1996), IEEE Computer Society, 1996.
- [YaST93] J. Yang, E. M. Siekierska, D. R. F. Taylor. Concept Design and Knowledge Acquisition for the Thematic Map Design Advisory System (TM-DAS). In *The Canadian Conf. on GIS '93*, 475–484, Ottawa, 1993.
- [Zade65] Zadeh, L. A. Fuzzy sets. *Information and Control*, Vol. 8, 1965, 338–353.
- [ZhBu95] F. B. Zhang, B. P. Battenfield. Object-oriented knowledge-based symbol selection for visualizing statistical information. *Int. Journal of Geographical Information Systems*, Vol. 9, No. 3, 1995, 293–315.